

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
14. Juni 2001 (14.06.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/42493 A2

- (51) Internationale Patentklassifikation⁷: C12Q 1/68 (81) Bestimmungsstaaten (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (21) Internationales Aktenzeichen: PCT/DE00/04381 (84) Bestimmungsstaaten (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (22) Internationales Anmeldedatum:
6. Dezember 2000 (06.12.2000)
- (25) Einreichungssprache:
Deutsch
- (26) Veröffentlichungssprache:
Deutsch
- (30) Angaben zur Priorität:
199 59 691.3 6. Dezember 1999 (06.12.1999) DE
- (71) Anmelder (*für alle Bestimmungsstaaten mit Ausnahme von US*): EPIGENOMICS AG [DE/DE]; Kastanienallee 24, D-10435 Berlin (DE).
- (72) Erfinder; und
- (75) Erfinder/Anmelder (*nur für US*): OLEK, Alexander [DE/DE]; Kyffhäuserstrasse 20, D-10781 Berlin (DE). PIEPENBROCK, Christian [DE/DE]; Schwartzkopffstrasse 7 B, D-10115 Berlin (DE).
- (74) Anwalt: SCHUBERT, Clemens; Joachimstrasse 9, D-10119 Berlin (DE).

Veröffentlicht:

— Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(54) Title: METHOD FOR THE PARALLEL DETECTION OF THE DEGREE OF METHYLATION OF GENOMIC DNA

(54) Bezeichnung: VERFAHREN ZUR PARALLELEN DETEKTION DES METHYLIERUNGSZUSTANDES VON GENOMISCHER DNA

(57) Abstract: The invention relates to a method for the parallel detection of the degree of methylation of genomic DNA wherein the following steps are performed: (a) chemical treatment at the 5' position of non-methylated cytosine bases converts said bases into uracil, thymidine or another base which exhibits hybridization behavior different to that of cytosine in a genomic DNA sample; (b) more than ten different fragments, each having less than 2000 base pairs in said chemically treated genomic DNA sample, are amplified simultaneously using synthetic oligonucleotides as a primer, whereby said primers each contain genomic sequences which are involved in gene regulation and/or transcribed and/or translated, such as those sequences which should be obtained after execution of steps (a); (c) the sequence contexts of all or a portion of the CpG dinucleotides or CpNpG trinucleotides contained in the amplified fragments are determined.

WO 01/42493 A2

(57) Zusammenfassung: Beschrieben ist ein Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA bei dem man folgende Schritte ausführt: (a) in einer genomischen DNA Probe wandelt man durch chemische Behandlung an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um; (b) aus dieser chemisch behandelten genomischen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind, gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer, wobei diese Primer jeweils Sequenzen aus an der Genregulation beteiligten und/oder transkribierten und/oder translatierten genomischen Sequenzen enthalten, wie sie nach einer Behandlung gemäß Schritt (a) vorliegen würden; (c) man bestimmt den Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.

Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA

Die vorliegende Erfindung betrifft ein Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA.

Die nach den methodischen Entwicklungen der letzten Jahre in der Molekularbiologie gut studierten Beobachtungsebenen sind die Gene selbst, die Übersetzung dieser Gene in RNA und die daraus entstehenden Proteine. Wann im Laufe der Entwicklung eines Individuums welches Gen angeschaltet wird und wie Aktivieren und Inhibieren bestimmter Gene in bestimmten Zellen und Geweben gesteuert wird, ist mit hoher Wahrscheinlichkeit mit Ausmaß und Charakter der Methylierung der Gene bzw. des Genoms korrelierbar. Insofern ist die Annahme naheliegend, daß pathogene Zustände sich in einem veränderten Methylierungsmuster einzelner Gene oder des Genoms äußern.

Stand der Technik sind Verfahren, welche das Studium von Methylierungsmustern einzelner Gene gestatten. Jüngere Fortentwicklungen dieser Methode erlauben auch die Analyse kleinsten Mengen an Ausgangsmaterial. Die vorliegende Erfindung beschreibt ein Verfahren zur parallelen Detektion des Methylierungszustandes genomicscher DNA Proben, wobei ausgehend von einer Probe gleichzeitig zahlreiche verschiedene Fragmente aus an der Genregulation beteiligten oder/und transkribierten und/oder translatierten Sequenzen amplifiziert werden und anschließend der Sequenzkontext in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide untersucht wird.

5-Methylcytosin ist die häufigste kovalent modifizierte Base in der DNA eukaryotischer Zellen. Sie spielt beispielsweise eine Rolle in der Regulation der Transkription, genomischem Imprinting und in der Tumorgenese. Die Identifizierung von 5-Methylcytosin als Bestandteil genetischer Information ist daher von erheblichem Interesse. 5-Methylcytosin-Positionen können jedoch nicht durch Sequenzierung identifiziert werden, da 5-Methylcytosin das gleiche Basenpaarungsverhalten aufweist

wie Cytosin. Darüber hinaus geht bei einer PCR-Amplifikation die epigenetische Information, welche die 5-Methylcytosine tragen, vollständig verloren.

Die Modifikation der genomischen Base Cytosin zu 5'-Methylcytosin stellt den bis heute wichtigsten und best untersuchten epigenetischen Parameter dar. Trotzdem gibt es bis heute zwar Methoden, umfassende Genotypen von Zellen und Individuen zu ermitteln, aber noch keine vergleichbaren Ansätze auch in großem Maße epigenotypische Information zu generieren und auszuwerten.

Es sind im Prinzip drei prinzipiell verschiedene Methoden bekannt, den 5-Methyl-Status eines Cytosins im Sequenzkontext zu bestimmen.

Die erste prinzipielle Methode beruht auf der Verwendung von Restriktionsendonukleasen (RE), welche „methylierungssensitiv“ sind. REs zeichnen sich dadurch aus, daß sie an einer bestimmten DNA-Sequenz, meist zwischen 4 und 8 Basen lang, einen Schnitt in die DNA einführen. Die Position solcher Schnitte kann dann durch Gelektrophorese, Transfer auf eine Membran und Hybridisierung nachgewiesen werden. Methylierungssensitiv bedeutet, daß bestimmte Basen innerhalb der Erkennungssequenz unmethyliert vorliegen müssen, damit der Schnitt erfolgen kann. Das Bandenmuster nach einem Restriktionsschnitt und Gelektrophorese ändert sich also je nach Methylierungsmuster der DNA. Allerdings befinden sich die wenigsten methylierbaren CpG innerhalb von Erkennungssequenzen von REs, können also nach dieser Methode nicht untersucht werden.

Die Empfindlichkeit dieser Methoden ist extrem niedrig (Bird, A.P., and Southern, E.M., J.Mol. Biol. 118, 27-47). Eine Variante kombiniert PCR mit dieser Methode, eine Amplifikation durch zwei auf beiden Seiten der Erkennungssequenz liegende Primer erfolgt nach einem Schnitt nur dann, wenn die Erkennungssequenz methyliert vorliegt. Die Empfindlichkeit steigt in diesem Fall auf theoretisch ein einziges Molekül der Zielsequenz, allerdings können mit hohem Aufwand nur einzelne Positionen untersucht werden (Shemer, R. et al., PNAS 93, 6371-6376). Wiederum ist Voraussetzung, daß sich die methylierbare Position innerhalb der Erkennungssequenz einer RE befindet.

Die zweite Variante beruht auf partieller chemischer Spaltung von Gesamt-DNA, nach dem Vorbild einer Maxam-Gilbert Sequenziereaktion, Ligation von Adaptoren an die so generierten Enden, Amplifikation mit generischen Primern und Auftrennung auf einer Gelektrophorese. Mit diesem Verfahren können definierte Bereiche bis zur Größe von weniger als tausend Basenpaaren untersucht werden. Das Verfahren ist allerdings so kompliziert und unzuverlässig, daß es praktisch nicht mehr verwendet wird (Ward, C. et al., J. Biol. Chem. 265, 3030-3033).

Eine relativ neue und die mittlerweile am häufigsten angewandte Methode zur Untersuchung von DNA auf 5-Methylcytosin beruht auf der spezifischen Reaktion von Bisulphit mit Cytosin, das nach anschließender alkalischer Hydrolyse in Uracil umgewandelt wird, welches in seinem Basen-Paarungsverhalten dem Thymidin entspricht. 5-Methylcytosin wird dagegen unter diesen Bedingungen nicht modifiziert. Damit wird die ursprüngliche DNA so umgewandelt, daß Methylcytosin, welches ursprünglich durch sein Hybridisierungsverhalten vom Cytosin nicht unterschieden werden kann, jetzt durch „normale“ molekularbiologische Techniken als einzig verbliebenes Cytosin beispielsweise durch Amplifikation und Hybridisierung oder Sequenzierung nachgewiesen werden kann. Alle diese Techniken beruhen auf Basenpaarung, welche jetzt voll ausgenutzt werden kann. Der Stand der Technik, was die Empfindlichkeit betrifft, wird durch ein Verfahren definiert, welches die zu untersuchende DNA in einer Agarose-Matrix einschließt, dadurch die Diffusion und Renaturierung der DNA (Bisulphit reagiert nur an einzelsträngiger DNA) verhindert und alle Fällungs- und Reinigungsschritte durch schnelle Dialyse ersetzt (Olek, A. et al., Nucl. Acids. Res. 24, 5064-5066). Mit dieser Methode können einzelne Zellen untersucht werden, was das Potential der Methode veranschaulicht. Allerdings werden bisher nur einzelne Regionen bis etwa 3000 Basenpaare Länge untersucht, eine globale Untersuchung von Zellen auf Tausende von möglichen Methylierungereignissen ist nicht möglich. Allerdings kann auch dieses Verfahren keine sehr kleinen Fragmente aus geringen Probenmengen zuverlässig analysieren. Diese gehen trotz Diffusionsschutz durch die Matrix verloren.

Eine Übersicht über die weiteren bekannte Möglichkeiten, 5-Methylcytosine

nachzuweisen, kann auch dem folgenden Übersichtsartikel entnommen werden: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 26, 2255 (1998).

Die Bisulphit-Technik wird bisher bis auf wenige Ausnahmen (z.B. Zeschnick, M. et al., Eur. J. Hum. Gen. 5, 94-98; Kubota T. et al., Nat. Genet. 16, 16-17) nur in der Forschung angewendet. Immer aber werden kurze, spezifische Stücke eines bekannten Gens nach einer Bisulphit-Behandlung amplifiziert und entweder komplett sequenziert (Olek, A. and Walter, J., Nat. Genet. 17, 275-276) oder einzelne Cytosin-Positionen durch eine „Primer-Extension-Reaktion“ (Gonzalgo, M. L. and Jones, P.A., Nucl. Acids. Res. 25, 2529-2531) oder Enzymschnitt (Xiong, Z. and Laird, P.W., Nucl. Acids. Res. 25, 2532-2534) nachgewiesen. Zudem ist auch der Nachweis durch Hybridisierung beschrieben worden (Olek et al, WO9928498).

Gemeinsamkeiten zwischen Promotoren bestehen nicht nur im Vorkommen von TATA- oder GC-Boxen sondern auch darin, für welche Transkriptionsfaktoren sie Bindestellen besitzen und in welchem Abstand diese sich zueinander befinden. Die existierenden Bindestellen für ein bestimmtes Protein stimmen in ihrer Sequenz nicht vollständig überein, es finden sich aber konservierte Folgen von mindestens 4 Basen, die durch das Einfügen von „Wobbles“, d. h. Positionen, an denen sich jeweils unterschiedliche Basen befinden, noch verlängert werden können. Des weiteren liegen diese Bindestellen in bestimmten Abständen zueinander vor.

Die Verteilung der DNA im Interphase-Chromatin, das den größten Teil des nuklearen Volumens einnimmt, unterliegt jedoch einer ganz speziellen Ordnung. So ist die DNA an mehreren Stellen an die nukleare Matrix, eine filamentöse Struktur an der Innenseite der nuklearen Membran, angeheftet. Diese Regionen bezeichnet man als matrix attachment regions (MAR) oder scaffold attachment regions (SAR). Das Anheften hat wesentlichen Einfluß auf die Transkription bzw. die Replikation. Diese MAR-Fragmente weisen keine konservativen Sequenzen auf, bestehen allerdings zu 70% aus A bzw. T und liegen in der Nähe von cis-agierenden Regionen, die die Transkription allgemein regulieren, und Topoisomerase II-Erkennungsstellen.

Neben Promotoren und Enhancern existieren weitere regulatorische Elemente für verschiedene Gene, sogenannte Insulators. Diese Insulators können z.B. die Wirkung des Enhancers auf den Promotor inhibieren, wenn sie zwischen Enhancer und Promotor liegen, oder aber, zwischen Heterochromatin und einem Gen gelegen, das aktive Gen vor dem Einfluß des Heterochromatins schützen. Beispiele für solche Insulators sind: 1. sogenannte LCR (locus control regions), welche aus mehreren gegenüber DNAase I hypersensitiven Stellen besteht; 2. bestimmte Sequenzen wie SCS (specialized chromatin structures) bzw. SCS', 350 bzw. 200 bp lang und hochresistent gegen Degradierung durch DNAase I und auf beiden Seiten von hypersensitiven Stellen flankiert (Abstand je 100 bp). An scs' bindet das Protein BEAF-32. Diese Insulators können auf beiden Seiten des Gens liegen.

Eine Übersicht über den Stand der Technik in der Oligomer Array Herstellung läßt sich auch einer im Januar 1999 erschienenen Sonderausgabe von Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999) und der dort zitierten Literatur entnehmen.

Patente, die sich allgemein auf die Verwendung von Oligomer Arrays und photolithographisches Maskendesign beziehen, sind z. B. US-A 5,837,832, US-A 5,856,174, WO-A 98/27430 und US-A 5,856,101. Zudem existieren einige Stoff- und Verfahrenspatente, welche die Verwendung photolabiler Schutzgruppen an Nukleosiden einschränken, so z. B. WO-A98/39348 und US-A 5,763,599.

Matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI) ist eine neue, sehr leistungsfähige Entwicklung für die Analyse von Biomolekülen (Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10.000 daltons. Anal. Chem. 60: 2299-2301). Ein Analytmolekül wird in eine im UV absorbierende Matrix eingebettet. Durch einen kurzen Laserpuls wird die Matrix ins Vakuum verdampft und das Analyt so unfragmentiert in die Gasphase befördert. Eine angelegte Spannung beschleunigt die Ionen in ein feldfreies Flugrohr. Auf Grund ihrer verschiedenen Massen werden Ionen unterschiedlich stark beschleunigt. Kleinere Ionen erreichen den Detektor früher als größere und die Flugzeit wird in die Masse der

Ionen umgerechnet.

Für die Abtastung eines immobilisierten DNA-Arrays sind vielfach fluoreszent markierte Sonden verwendet worden. Besonders geeignet sind für die Fluoreszenzmarkierung ist das einfache Anbringen von Cy3 und Cy5 Farbstoffen am 5'OH der jeweiligen Sonde. Die Detektion der Fluoreszenz der hybridisierten Sonden erfolgt beispielsweise über ein Konfokalmikroskop. Die Farbstoffe Cy3 und Cy5 sind, neben vielen anderen, kommerziell erhältlich.

Um die erwartete Anzahl von amplifizierten Fragmenten ausgehend von einer beliebigen Templat-DNA und zweien nicht für jeweils eine bestimmte Position spezifischen Primern zu berechnen, muß ein statistisches Modell über den Aufbau des Genoms zu Grunde gelegt werden.

Wir geben hier die Berechnung für drei Modelle an, beziehen uns allerdings in diesem Patent auf die in Modell 3 beschriebene Methode.

Modell 1:

Im einfachsten Fall wird angenommen, daß ein primärer DNA-Strang eine Zufallsfolge von vier gleich häufig vorkommenden Basen ist. Damit ergibt sich als Wahrscheinlichkeit, daß sich für einen beliebigen Primer *PrimA* (der Länge k) an einer gegebenen Stelle im Genom eine perfekte Basenpaarung ergibt:

$$P_a(\text{PrimA}) = 0.25^k \quad (\text{Modell 1 für DNA})$$

(diese Wahrscheinlichkeit ist für den sense- und anti-sense-Strang der DNA gleich)

Bei einer Bisulfitbehandlung der DNA werden diejenigen Cytosine durch Uracil ersetzt, die nicht zu einem methylierten CG gehören. Das Basenpaarungsverhalten des Uracils entspricht dem des Thymins. Da CG in der DNA sehr selten sind (unter zwei Prozent), kann die statistische Häufigkeit der Cs nach der Bisulfitbehandlung vernachlässigt werden. Die Wahrscheinlichkeit, daß sich für einen Primer *PrimB* (Länge k , davon a As,

t Ts, *g* Gs und *c* Cs) auf bisulfitbehandelter DNA eine perfekte Basenpaarung ergibt, ist unterschiedlich für einen mit Bisulfit behandelten Strang und den zugehörigen anti-sense Strang:

$$P_{Is}(PrimB) = 0.5^a * 0.25^t * 0.25^c * 0^g \quad (\text{Modell 1 für Bisulfit-DNA-Strang})$$

$$P_{Ia}(PrimB) = 0.25^a * 0.5^t * 0^c * 0.25^g \quad (\text{Modell 1 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

(wenn der Primer C oder G enthält, wird somit einer der Wahrscheinlichkeitswerte 0).

Modell 2:

Zählungen der Basenhäufigkeiten der DNA ergeben, daß die vier Basen in der DNA nicht gleichverteilt sind. Entsprechend kann man aus DNA-Datenbanken folgende Häufigkeiten (Wahrscheinlichkeiten für ein Vorkommen) der Basen ermitteln.

$$P_{DNA}(A) = 0.2811$$

$$P_{DNA}(T) = 0.2784$$

$$P_{DNA}(C) = 0.2206$$

$$P_{DNA}(G) = 0.2199$$

Als Grundlage für diese Statistik (und die folgenden für Modell 2 und 3) dienen ca. 6% des Genoms vom Homo Sapiens aus High Throughput Sequencing Projekten (Datenbank "htgs" vom NIH/NCBI vom 6.9.1999). Die Gesamtmenge der Daten beträgt mehr als 1.5×10^8 Basenpaare, was einem Schätzfehler für die Einzelwahrscheinlichkeiten kleiner 10^{-5} entspricht.

Mit Hilfe dieser Werte läßt sich das Modell 1 verbessern.

Damit ist die Wahrscheinlichkeit, daß sich für einen Primer *PrimC* (Länge *k*, davon *a* As, *t* Ts, *g* Gs und *c* Cs) eine perfekte Basenpaarung ergibt:

$$P_2(PrimC) = P_{DNA}(T)^a * P_{DNA}(A)' * P_{DNA}(C)^g * P_{DNA}(G)^c \quad (\text{Modell 3 für DNA})$$

Für den mit Bisulfit behandelten Strang ergeben sich folgende Wahrscheinlichkeiten unter der Annahme, daß alle CpG-Positionen methyliert sind (man erhält eine gleiche Statistik für die Bisulfitbehandlung des DNA-sense- und des DNA-antisense-Stranges):

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

Damit ergibt sich als Wahrscheinlichkeit, daß sich für einen Primer *PrimD* (Länge k , davon a As, t Ts, g Gs und c Cs) eine perfekte Basenpaarung ergibt:

$$P_{2s}(PrimD) = P_{bDNA}(T)^a * P_{bDNA}(A)' * P_{bDNA}(C)^g * P_{bDNA}(G)^c \quad (\text{Modell 3 für Bisulfit-DNA-Strang})$$

$$P_{2a}(PrimD) = P_{bDNA}(A)^a * P_{bDNA}(T)' * P_{bDNA}(G)^g * P_{bDNA}(C)^c \quad (\text{Modell 3 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

Modell 3:

Wesentliche Schätzfehler in Modell 2 ergeben sich vor allem bei der mit Bisulfit behandelten DNA aus der Tatsache, daß C nur noch im Kontext CG auftreten kann. Modell 3 berücksichtigt diese Eigenschaft und nimmt an, daß die primäre DNA eine Zufallsfolge mit Abhängigkeit direkt benachbarter Basen ist (Markov-Kette erster Ordnung). Die empirisch aus der Datenbank (vollständig methyliert; mit Bisulfat behandelt) ermittelten paarweisen Basenwahrscheinlichkeiten ergeben sich gleich für beide DNA-Stränge als $P_{bDNA}(\text{von}, \text{nach})$ aus der folgenden Tabelle:

Von\nach	A	C	G	T
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge) $P_{rbDNA}(von; nach)$

Von\nach	A	C	G	T
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A) = 0.4850 ;$$

$$P_{rbDNA}(C) = 0.2199$$

$$P_{rbDNA}(G) = 0.0140$$

$$P_{rbDNA}(T) = 0.2811$$

Damit hängt die Wahrscheinlichkeit, daß sich für einen Primer *PrimE* (mit der Basenfolge $B_1 B_2 B_3 B_4 \dots$; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots \quad (\text{Modell 3 für})$$

Bisulfit-DNA-Strang)

$$P_{3u}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots \quad (\text{Modell 3 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

Berechnung der Anzahl der zu erwartenden amplifizierten Fragmente

Die mit Bisulfit behandelte DNA wird unter Benutzung einer Anzahl Primer amplifiziert. Aus Sicht des Modells besteht die DNA aus je einem sense- und einem anti-sense-Strang der Länge N Basen (alle Chromosomen werden hier zusammengefaßt). Für einen Primer $Prim$ ist zu erwarten, daß er auf dem sense-Strang

$$N * P_s(Prim)$$

perfekte Basenpaarungen ergibt - für diese Berechnung können die Funktionen P_{1s} , P_{2s} oder P_{3s} von Modell 1, 2 oder 3 eingesetzt werden, je nach gewünschter Abschätzungsgüte. Werden mehrere Primer ($PrimU$, $PrimV$, $PrimW$, $PrimX$, etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) &= P_s(PrimU) \\ &+ (1 - P_s(PrimU)) P_s(PrimV) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV)) P_s(PrimW) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW)) P_s(PrimX) \\ &+ \dots \end{aligned}$$

Und damit als Anzahl der zu erwartenden perfekten Basenparungen mit irgendeinem der Primer

$$N * P_s(Primers)$$

Für die Bestimmung von $P_a(Primers)$ auf dem anti-sense-Strang werden die analogen Gleichungen verwendet. Ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge M ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet. Die Wahrscheinlichkeit dafür ist

$$P_a(Primers) \sum_{i=0}^{M-2} (1 - P_a(Primers))$$

Für große M und kleine $P_a(Primers)$ kann dieses durch folgenden Ausdruck berechnet werden:

$$\frac{1 - P_a(Primers)}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1]$$

Für die Gesamtzahl F der Fragmente, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit:

$$F = N * P_s(Primers) \frac{(1 - P_a(Primers))}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] + N * P_a(Primers) \frac{(1 - P_s(Primers))}{\log(1 - P_s(Primers))} [(1 - P_s(Primers))^M - 1]$$

Diese Methode liefert einen präzisen Erwartungswert für die Vorhersage der Anzahl der Bindungssites bestimmter Sequenzen an ein beliebiges zuvor mit Bisulfit behandeltes genomicsches DNA Fragment. Sie dient hier als Grundlage für die Berechnung der statistisch erwarteten Anzahl von Amplifikaten in einer PCR-Reaktion ausgehend von zwei Primersequenzen und einer DNA der Länge N , wobei nur die Amplifikate berücksichtigt werden, die eine Anzahl von M Nukleotiden nicht überschreiten. In diesem Patent wird davon ausgegangen, daß M den Wert 2000 hat.

Die bekannten Verfahren für den Nachweis von Cytosin Methylierungen in genomischer DNA sind prinzipiell nicht so ausgelegt, daß eine Vielzahl von Zielregionen im zu untersuchenden Genom gleichzeitig erfaßt werden. Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zu schaffen, mit dem eine Probe genomischer DNA gleichzeitig an mehreren Positionen gleichzeitig auf Cytosin Methylierung hin untersucht werden kann.

Die Aufgabe wird durch die kennzeichnenden Merkmale des Anspruchs 1 gelöst. Vorteilhafte Weiterbildungen der Merkmal sind in den abhängigen Ansprüchen gekennzeichnet.

Im Unterschied zu anderen Verfahren kann nach chemischer Vorbehandlung der DNA durch Verwendung entsprechend angepaßter Primerpaare eine Amplifikation von vielen Zielregionen gleichzeitig erfolgen. Dabei ist es nicht unbedingt notwendig den Sequenzkontext aller dieser Zielregionen vorab zu kennen, da in vielen Fällen, wie nachfolgend auch beispielhaft aufgeführt, Konsensussequenzen aus der Sequenzierung verwandter Zielregionen bekannt sind, die wie unten beschrieben für das Design für bestimmte Zielregionen spezifischer oder selektiver Primerpaare eingesetzt werden können. Das Verfahren ist dann erfolgreich angewandt, wenn die Amplifikation der chemisch vorbehandelten genomischen DNA mehr Fragmente bis maximal 2000 Basenpaare Länge als statistisch zu erwarten aus den jeweils zu untersuchenden Zielregionen liefert.

Dabei wird der statistische Erwartungswert für die Anzahl dieser Fragmente über die im Stand der Technik aufgeführten Formeln berechnet. Die Anzahl der im Amplifikationsschritt hergestellten Fragmente kann dagegen mittels einer beliebigen molekularbiologischen, chemischen oder physikalischen Methode nachgewiesen werden.

Für die Durchführung der erforderlichen statistischen Betrachtungen, die auch für die unten aufgeführten Ansprüche relevant sind, werden die folgenden Werte angenommen:

Das menschliche haploide Genom enthält 3 Milliarden Basenpaare und 100.000 Gene, die wiederum im Mittel eine 2000 Basenpaare lange mRNA codieren, die Gene inklusive der Introns sind durchschnittlich 15000 Basenpaare lang. Promotoren umfassen je Gen 1000 Basenpaare durchschnittlich. Ist daher der statistische Erwartungswert für die Anzahl der Amplifikate, die ausgehend von zwei Primern in transkribierten Sequenzen liegen, zu berechnen, so ist zunächst der Erwartungswert für das Gesamtgenom nach obiger Formel (Methode 3) zu berechnen und mit dem Anteil der transkribierten Sequenzen am Gesamtgenom zu berechnen. Analog wird für Teile eines beliebigen Genoms sowie für Promotoren und translatierte Sequenzen (mRNA codierend) vorgegangen.

Die vorliegende Erfindung beschreibt somit ein Verfahren zur parallelen Detektion des Methylierungszustandes genetischer DNA. Dabei sollen mehrere Cytosin-Methylierungen in einer DNA-Probe gleichzeitig analysiert werden. Dazu werden die folgenden Verfahrensschritte nacheinander ausgeführt:

Zuerst wird eine genetische DNA Probe derart chemisch behandelt, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden. Bevorzugt wird dazu die oben beschriebene Behandlung genetischer DNA mit Bisulfit (Hydrogensulfat, Disulfat) und anschließender alkalischer Hydrolyse verwendet, die zu einer Umwandlung nicht methylierter Cytosin-Nukleobasen in Uracil führt.

In einem zweiten Verfahrensschritt werden aus der vorbehandelten genetischen DNA mehr als zehn unterschiedliche Fragmente gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer amplifiziert, wobei mehr als doppelt so viele Fragmente als statistisch zu erwarten aus an der Genregulation beteiligten, transkribierten und/oder translatierten Sequenzen stammen. Dies kann mittels verschiedener Methoden erreicht werden.

In einer bevorzugten Variante des Verfahrens enthält mindestens eines der für die Amplifikation verwendeten Oligonukleotide weniger Nukleobasen als es statistisch für

eine sequenzspezifische Hybridisierung an die chemisch behandelte genomische DNA Probe erforderlich wäre, was zur Amplifikation mehrerer Fragmente gleichzeitig führen kann. Dabei ist die Gesamtzahl der in diesem Oligonukleotid enthaltenen Nukleobasen kleiner als 17. In einer besonders bevorzugten Variante des Verfahrens ist die Anzahl der in diesem Oligonukleotid enthaltenen Nukleobasen kleiner als 14.

In einer weiteren, bevorzugten Variante des Verfahrens werden für die Amplifikation mehr als 4 Oligonukleotide mit unterschiedlicher Sequenz gleichzeitig in einem Reaktionsgefäß verwendet. In einer besonders bevorzugten Varianten werden zur Herstellung eines komplexen Amplifikates mehr als 26 verschiedene Oligonukleotide gleichzeitig verwendet. In einer besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl, wie statistisch zu erwarten, aus an der Regulation von Genen beteiligten Genomabschnitten, z.B. Promotoren und Enhancern, stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre. In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, die in mindestens einer Zelle des jeweiligen Organismus in mRNA transkribiert werden, oder aber aus nach der Transkription in mRNA gespliceten Genomabschnitten (Exons), als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, welche für Teile einer oder mehrerer Genfamilien kodieren, oder aber sie stammen aus Genomabschnitten, welche für sogenannte „matrix attachment sites“ (MARs)-charakteristische Sequenzen enthalten, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, welche als sogenannte „boundary elements“ die Verpackungsdichte des Chromatins organisieren, oder aber sie stammen aus multiple drug resistance gene“ (MDR)-

Promotoren oder kodierenden Regionen, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren, besonders bevorzugten Variante des Verfahrens werden zur Amplifikation der beschriebenen Fragmente zwei Oligonukleotide oder zwei Klassen von Oligonukleotiden verwendet, von denen eines oder eine Klasse außer im Kontext CpG oder CpNpG zwar die Base C enthalten kann, nicht aber die Base G und von denen das andere oder die andere Klasse außer im Kontext CpG oder CpNpG zwar die Base G, nicht aber die Base C enthalten kann.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide durchgeführt, von denen eines eine vier bis sechzehn Basen lange Sequenz enthält, die zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, an welches einer der Faktoren

AhR/Arnt	aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator
Arnt	aryl hydrocarbon receptor nuclear translocator
AML-1a	CBFA2; core-binding factor, runt domain, alpha subunit 2 (acute myeloid leukemia 1; aml1 oncogene)
AP-1	activator protein-1 (AP-1); Synonyme: c-Jun
C/EBP	CCAAT/enhancer binding protein
C/EBPalpha	CCAAT/enhancer binding protein (C/EBP), alpha
C/EPbeta	CCAAT/enhancer binding protein (C/EBP), beta
CDP	CUTL1; cut (<i>Drosophila</i>)-like 1 (CCAAT displacement protein)
CDP	CUTL1; cut (<i>Drosophila</i>)-like 1 (CCAAT displacement protein)
CDP CR1	complement component (3b/4b) receptor 1
CDP CR3	complement component (3b/4b) receptor 3
CHOP-C/EBPalpha	DDIT; DNA-damage-inducible transcript 3/CCAAT/enhancer binding protein (C/EBP), alpha
c-Myc/Max	avian myelocytomatis viral oncogene/MYC-ASSOCIATED FACTOR X
CREB	cAMP responsive element binding protein
CRE-BP1	CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2, CREB2, CREBP1; now ATF2; activating transcription factor 2
CRE-BP1/c-Jun	activator protein-1 (AP-1); Synonyme: c-Jun

CREB	MP responsive element binding protein
E2F	E2F transcription factor (originally identified as a DNA-binding protein essential E1A-dependent activation of the adenovirus E2 promoter)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
Egr-1	early growth response 1
Egr-2	early growth response 2 (Krox-20 (Drosophila) homolog)
ELK-1	ELK1, member of ETS (environmental tobacco smoke) oncogene family
Freac-2	FKHL6; forkhead (Drosophila)-like 6; FORKHEAD-RELATED ACTIVATOR 2; FREAC2
Freac-3	FKHL7; forkhead (Drosophila)-like 7; FORKHEAD-RELATED ACTIVATOR 3; FREAC3
Freac-4	FKHL8; forkhead (Drosophila)-like 8; FORKHEAD-RELATED ACTIVATOR 4; FREAC4
Freac-7	FKHL11; forkhead (Drosophila)-like 9; FORKHEAD-RELATED ACTIVATOR 7; FREAC7
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-2	GATA-binding protein 2/Enhancer-Binding Protein GATA2
GATA-3	GATA-binding protein 3/Enhancer-Binding Protein GATA3
GATA-X	
HFH-3	FKHL10; forkhead (Drosophila)-like 10; FORKHEAD-RELATED ACTIVATOR 6; FREAC6
HNF-1	TCF1; transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
HNF-4	hepatocyte nuclear factor 4
IRF-1	interferon regulatory factor 1
ISRE	interferon-stimulated response element
Lmo2 complex	LIM domain only 2 (rhombotin-like 1)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
myogenin/NF-1	Myogenin (myogenic factor 4)/Neurofibromin 1; NEUROFIBROMATOSIS, TYPE I
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
NF-E2	NFE2; nuclear factor (erythroid-derived 2), 45kD
NF-kappaB (p50)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p50 subunit
NF-kappaB (p65)	nuclear factor of kappa light polypeptide gene enhancer in B-

	cells p65 subunit
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NRSF	NEURON RESTRICTIVE SILENCER FACTOR; REST; RE1-silencing transcription factor
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
P300	E1A (adenovirus E1A oncprotein)-BINDING PROTEIN, 300-KD
P53	tumor protein p53 (Li-Fraumeni syndrome); TP53
Pax-1	paired box gene 1
Pax-3	paired box gene 3 (Waardenburg syndrome 1)
Pax-6	paired box gene 6 (aniridia, keratitis)
Pbx 1b	pre-B-cell leukemia transcription factor
Pbx-1	pre-B-cell leukemia transcription factor 1
RORalpha2	RAR-RELATED ORPHAN RECEPTOR ALPHA; RETINOIC ACID-BINDING RECEPTOR ALPHA
RREB-1	ras responsive element binding protein 1
SP1	simian-virus-40-protein-1
SP1	simian-virus-40-protein-1
SREBP-1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
SRY	sex determining region Y
STAT3	signal transducer and activator of transcription 1, 91kD
Tal-1alpha/E47	T-cell acute lymphocytic leukemia 1/transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TATA	cellular and viral TATA box elements
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
TCF11/MafG	v-maf musculoaponeurotic fibrosarcoma (avian) oncogene family, protein G
TCF11	Transcription Factor 11; TCF11; NFE2L1; nuclear factor (erythroid-derived 2)-like 1
USF	upstream stimulating factor
Whn	winged-helix nude

X-BP-1
YY1

X-box binding protein 1 oder
ubiquitously distributed transcription factor belonging to
the GLI-Kruppel class of zinc finger proteins

bindet, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide oder zweier Klassen von Oligonukleotiden durchgeführt, von denen eines oder die eine Klasse die vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide oder zweier Klassen von Oligonukleotiden durchgeführt, von denen eines oder die eine Klasse eine der Sequenzen

TCGCGTGTA, TACACGCGA, TGTACGCGA, TCGCGTACA,
TTGCGTGTT, AACACGCAA, GGTACGTAA, TTACGTACC,
TCGCGTGTT, AACACGCGA, GGTACGCGA, TCGCGTACC,
TTGCGTGTA, TACACGCAA, TGTACGTAA, TTACGTACA,
TACGTG, CACGTA, TACGTG, CACGTA,

ATTGCGTGT, ACACCGCAAT, GTACGTAAT, ATTACGTAC,
ATTGCGTGTA, TCACCGCAAT, TTACGTAAT, ATTACGTAA,
ATCGCGTGA, TCACCGCGAT, TTACCGCGAT, ATCGCGTAA,
ATCGCGTGT, ACACCGCGAT, GTACCGCGAT, ATCGCGTAC,
TGTGGT, ACCACA, ATTATA, TATAAT,

TGAGTTAG, CTAACTCA, TTGATTAA, TAAATCAA,
TGATTTAG, CTAAATCA, TTGAGTTA, TAACTCAA,

TTTGGT, ACCAAA, ATTAAA, TTTAAT,
TGTGGA, TCCACA, TTTATA, TATAAA,
TTTGGG, TCCAAA, TTTAAA, TTTAAA,
TGTGGT, ACCACA, ATTATA, TATAAT,

ATTAT, ATAAT, GTAAT, ATTAC,
ATTGT, ACAAT, GTAAT, ATTAC,

GAAAG, CTTTC, TTTTT, AAAAAA,
GTAAT, ATTAC, ATTGT, ACAAT,
GAAAT, ATTTTC, ATTTTT, AAAAT,
GTAAG, CTTAC, TTTGT, ACAAAA,
TTAATAATCGAT, ATCGATTATTAA, ATCGATTATTGG, CCAATAATCGAT,
ATCGATTA, TAATCGAT, TAATCGAT, ATCGATTA,

ATCGATCGG, CCGATCGAT, TCGATCGAT, ATCGATCGA,
ATCGATCGT, ACGATCGAT, GCGATCGAT, ATCGATCGC,

TATCGATA, TATCGATA, TATCGGTG, CACCGATA,
TATTAATA, TATTAATA, TATTGGTG, CACCAATA,

GTGTAATATT, AAATATTACAC, GGGTATTGTAT, ATACAATACCC,
GTGTAATTTT, AAAAATTACAC, GGGGATTGTAT, ATACAATCCCC,
ATGTAATTTT, AAAAATTACAT, GGGGATTGTAT, ATACAATCCCC,
ATGTAATATT, AAATATTACAT, GGGTATTGTAT, ATACAATACCC,
ATTACGTGGT, ACCACGTAAT, ATTACGTGGT, ACCACGTAAT,
TGACGTAA, TTACGTCA, TTACGTTA, TAACGTAA,
TGACGTTA, TAACGTCA, TGACGTTA, TAACGTCA,
TTACGTAA, TTACGTAA, TTACGTAA, TTACGTAA,
TGACGTTA, TAACGTCA, TAACGTTA, TAACGTTA,

TGACGT, ACGTCA, GCGTTA, TAACGC,
TGACGT, ACGTCA, ACGTTA, TAACGT,
TTTCGCGT, ACGCGAAA, GCGCGAAA, TTTCGCGC,
TTTGGCGT, ACGCCAAA, GCGTTAAA, TTAAACGC,

TAGGTGTTA, TAACACCTA, TAATATTG, CAAATATTA,
TAGGTGTTT, AAACACCTA, GAATATTG, CAAATATTC,

GTAGGTGG, CCACCTAC, TTATTG, ACAAAATAA,
GTAGGTGT, ACACCTAC, ATATTG, ACAAAATAT,

TGCCTGGCGG, CCGCCCACGCA, TCGTTTACGTA, TACGTAAACGA,
TGCCTGGCGT, ACGCCCACGCA, ACGTTTACGTA, TACGTAAACGT,

TGCCTAGGCCT, ACGCCTACGCA, ACGTTTACGTA, TACGTAAACGT,
TGCCTAGGCCT, CCGCCTACGCA, TCGTTTACGTA, TACGTAAACGA,
ATAGGAAGT, ACTTCCTAT, ATTTTTGT, ACAAAAAAT,

TCGGAAGT, ACTTCCGA, ATTTTCGG, CCGAAAAT,
TCGGAAGT, ACTTCCGA, GTTTTCGG, CCGAAAAC,
TCGGAAAT, ATTTCCGA, ATTTTCGG, CCGAAAAT,
TCGGAAAT, ATTTCCGA, GTTTTCGG, CCGAAAAC,
GTAAATAA, TTATTTAC, TTGTTTAT, ATAAACAA,
GTAAATAAATA, TATTTATTTAC, TGTTTATTAT, ATAAATAAACA,
AAAGTAAATA, TATTTACTTT, TGTTTATTAT, AAAATAAACA,
AATGTAAATA, TATTTACATT, TGTTTATATT, AATATAAACA,
TAAGTAAATA, TATTTACTTA, TGTTTATTAT, TAAATAAACA,
TATGTAAATA, TATTTACATA, TGTTTATATA, TATATAAACA,
ATAAAATA, TATTTAT, TGTTTAT, ATAAACA,
ATAAAATA, TATTTAT, TATTTAT, ATAAATA,
GATA, TATC, TATT, AATA,
TAGATAA, TTATCTA, TTATTTG, CAAATAA,
TTGATAA, TTATCAA, TTATTAG, CTAATAA,
GATAA, TTATC, TTATT, AATAA,
GATG, CATC, TATT, AATA,
GATAG, CTATC, TTATT, AATAA,
GATAAG, CTTATC, TTTATT, AATAAA,
TGTTTATTAT, TAAATAAACA, TAAATAAATA, TATTTATTAT,
TGTTTGTAT, TAAACAAACA, TAAATAAATA, TATTTATTAT,
TATTTATTAT, TAAATAAATA, TAAATAAATA, TATTTATTAT,
TATTTGTAT, TAAACAAATA, TAAATAAATA, TATTTATTAT,
GTTAATGATT, AATCATTAAC, AATTATTAAT, ATTAATAATT,
GTTAATTATT, AATAATTAAC, AATAATTAAT, ATTAATTATT,
GTTAATTAAT, ATTAATTAAC, ATTAATTAAT, ATTAATTAAT,
GTTAATGAAT, ATTCACTAAC, ATTTATTAAT, ATTAATAAAAT,
TAAAGTTA, TAAACTTTA, TGAATTTG, CAAAATTCA,
TAAAGGTTA, TAACCTTTA, TGATTTTG, CAAAAATCA,
AAAGTGAATT, AATTCACCTT, GGTTTATTAT, AAAATAAAACC,
AAAGCGAAATT, AATTCGCTTT, GGTTTCGTTT, AAAACGAAACC,
TAGTTTATTATTTT, AAAAAAATAAAACTA, GGGAAAGTGAATTG,
CAATTCACCTTCCC,
TAGTTTATTATTTT, AAAAAAATAAAACTA, GGAAAAGTGAATTG,
CAATTCACCTTCC,
TAGTTTTTTTTTTT, AAAAAAAAAAAACTA, GGAAAAGAGAAATTG,
CAATTCCTCTTTCC,

TAGTTTTTTTTTTT, AAAAAAAAAGACTA, GGGAAAGAGAAATTG,
CAATTTCTCTTCCC,
TAGGTG, CACCTA, TATTTG, CAAATA,

TTTTAAAAATAATTTT, AAAATTATTTTAAAA, AGGGTTATTTTAGAG,
CTCTAAAAATAACCCT,
TTTTAAAAATAATTTT, AAAATTATTTTAAAA, GGAGTTATTTTAGAG,
CTCTAAAAATAACTCC,
TTTTAAAAATAATTTT, AAAATTATTTTAAAA, AGAGTTATTTTAGAG,
CTCTAAAAATAACTCT,
TTTTAAAAATAATTTT, AAAATTATTTTAAAA, GGGGTTATTTTAGAG,
CTCTAAAAATAACCCC,

TGTTATTAAAAATAGAAA, TTTCTATTTTAATAACA,
TTTTTATTTTAGTAATA, TATTACTAAAAATAAAAAA,
TGTTATTAAAAATAGAAT, ATTCTATTTTAATAACA,
GTTTATTTTAGTAATA, TATTACTAAAAATAAAC,
TTTGGTAT, ATACCAAA, GTGTTAAA, TTTAACAC
GGGA, TCCCC, TTTTT, AAAAA,

TAGGGG, CCCCTA, TTTTTA, TAAAAAA,
GAGGGG, CCCCTC, TTTTTT, AAAAAAA,
TGTTGAGTTAT, ATAACCTCAACA, ATGATTTAGTA, TACTAAATCAT,
TGTTGATTTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,
TGTTGAGTTAT, ATAACCTCAACA, ATGATTTAGTA, TACTAAATCAT,
TGTTGATTTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,
GGGGATTTTT, AAAAAATCCCC, GGGAAATTTTT, AAAAAATTCCC,
GGGGATTTTT, AAAAAATCCCC, GGGGATTTTT, AAAAAATCCCC,
GGGGATTTTT, AAAAAATCCCC, GGAAATTTTT, AAAAAATTCC,
GGGAATTTTT, AAAAAATTCCC, GGAAATTTTT, AAAAAATTCC,
GGGATTTTTT, AAAAAATCCC, GGAAAGTTTT, AAAACTTTCC,
GGGAATTTTT, AAAAAATTCCC, GGGAAATTTTT, AAAAAATTCCC,
GGGATTTTT, AAAAAATCCC, GGGAAAGTTTT, AAAACTTCCC,
GGGATTTTTA, TAAAAAAATCCC, TGGAAGTTTT, AAAACTTCCA,
TTTAGTATTACGGATAGAGGT, ACCTCTATCCGTAAATACTAAA,
TTTTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAAAAC,
TTTAGTATTACGGATAGAGTT, AACGCTATCCGTAAATACTAAA,
GGCGTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAACGCC,
TTTAGTATTACGGATAGCGGT, ACCGCTATCCGTAAATACTAAA,
GTCGTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAACGAC,

ATATGTAAAT, ATTTACATAT, ATTTGTATAT, ATATACAAAT,
TTATGTAAAT, ATTTACATAA, ATTTGTATAA, TTATACAAAT,
GAATATTAA, TAAATATTCA, TGAATATT, AAATATTCA,

GAATATGTA, TACATATTTC, TGTATATTT, AAATATACA,
ATAAT, ATTAT, ATTAT, ATAAT,
GTAAT, ATTAC, ATTAT, ATAAT,
AATGTAAAT, ATTTACATT, ATTTGTATT, AATACAAAT,
ATTTGTATATT, AATATACAAAT, GGTATGTAAAT, ATTTACATACC,
ATTTGTATATT, AATATACAAAT, AATATGTAAAT, ATTTACATATT,
ATTTGTATATT, AATATACAAAT, AGTATGTAAAT, ATTTACATACT,
ATTTGTATATT, AATATACAAAT, GATATGTAAAT, ATTTACATATC,
AGGAGT, ACTCCT, ATTTTT, AAAAAT,
GGGAGT, ACTCCC, ATTTTT, AAAAAT,
GGATATGTTCGGGTATGTTT, AACATACCCGAACATATCC,
GGATATGTTCGGGTATGTTT, AACATACCCGAACATATCC,
GGATATGTTCGGGTATGTTT, AACATACCCGAACATATCC,
AGATATGTTCGGGTATGTTT, AACATACCCGAACATATCC,
TCGTTTCGTTTAGATAT, ATATCTAAAACGAAACGA,
ATATTAGAGCGGAACGG, CCGTTCCGCTCTAAATAT,
CGTTACGGTT, AACCGTAACG, AATCGTGACG, CGTCACGATT,
CGTTACGGTT, AACCGTAACG, GATCGTGACG, CGTCACGATC,
CGTTACGTTT, AACCGTAACG, AAGCGTGACG, CGTCACGCTT,
CGTTACGTTT, AACCGTAACG, GAGCGTGACG, CGTCACGCTC,
TTTACGTATGA, TCATACGTAAA, TTATGCGTGAA, TTCACGCATAA,
TTTACGTTTGA, TCAAACGTAAA, TTAAGCGTGAA, TTCACGCTTAA,
TTTACGTTTTA, TAAAACGTAAA, TGAAGCGTGAA, TTCACGCTTCA,
TTTACGTATTA, TAATACGTAAA, TGATGCGTGAA, TTCACGCATCA,
AATTAATTAA, TTAATTAATT, TTGATTGATT, AATCAATCAA,
TATTAATTAA, TTAATTAATA, TTGATTGATG, CATCAATCAA,
TAATTAT, ATAATTA, ATGATTG, CAATCAT,
TAGGTTA, TAACCTA, TGATTTA, TAAATCA,
TTTAAATATTTT, AAAAATATTTAAAA, GGGGGTGGTTGGGG,
CCCCAAACACCCCC,
TTTAAATTTTTT, AAAATAATTTAAAA, GGGGTGGTTGGGG,
CCCCAAACACCCCC,
TTTAAATTTTTT, AAAAAAATTTAAAA, GGGGGGGTTGGGG,
CCCCAAACACCCCC,
TTTAAATAATTTT, AAAATTATTTAAAA, GGGGTTGGTTGGGG,
CCCCAAACACCCCC,
GAGGCAGGGG, CCCCGCCTC, TTTCGTTT, AAAACGAAA,

GAGGTAGGG, CCCTACCTC, TTTTGT₄, AAAACAAAAA,
AAGGC₄GGG, CCCCGCCTT, TTTCGT₄TT, AAAACGAAA,
AAGGTAGGG, CCCTACCTT, TTTTGT₄TT, AAAACAAAAA,

GGGGCGGGGT, ACCCGCCCCC, ATTCGTTTT, AAAAACGAAAT,
GGGGCGGGGT, ACCCGCCCCC, GTTCGTTTT, AAAAACGAAAC,
TATTATTTAT, ATAAAATAATA, GTGGGGTGATA, TATCACCCAC,
GATTATTTAT, ATAAAATAATC, GTGGGGTGATT, AATCACCCAC,

ATTACGTGAT, ATCACGTAAT, ATTACGTGAT, ATCACGTAAT,
ATTACGTGAT, ATCACGTAAT, GTTACGTGAT, ATCACGTAAC,

TTTATATGG, CCATATAAAA, TTATATAAGG, CCTTATATAA,
TTATATATGG, CCATATATAA, TTATATATGG, CCATATATAA,
AAATAAT, ATTATT, GTTGT₃, AAACAAC,
AAATTAA, TTAATT, TTAGTT, AAACTAA,
AAATTAT, ATAATT, GTAGTT, AAACTAC,
AAATAAA, TTTATT, TTTGT₃, AAACAAA,

ATTTTCGGAAATG, CATTTCGAAAAAT, TATTTTCGGAAAT,
ATTTCCGAAAATA,
ATTTTCGGAAATG, CATTTCGAAAAAT, TATTTTCGGAAAT,
ATTTCCGAAAATA,
ATTTTCGGAAATG, CATTCCGAAAAT, TATTTTCGGAAAT,
ATTTCCGAAAATA,
ATTTTCGGGAAGTG, CACTTCCGAAAAT, TATTTTCGGAAAT,
ATTTCCGAAAATA,

AATAGATGTT, AACATCTATT, AATATTGTT, AACAAATATT,
AATAGATGGT, ACCATCTATT, ATTATTGTT, AACAAATAAT,

GTATAAATA, TATTTATAC, TATTTATAT, ATATAAATA,
GTATAAATG, CATTTATAC, TATTTATAT, ATATAAATA,
GTATAAAA, TTTTATAC, TTTTATAT, ATATAAAA,
GTATAAAAG, CTTTATAC, TTTTATAT, ATATAAAA,
TTATAAATA, TATTTATAA, TATTTATAG, CTATAAATA,
TTATAAAATG, CATTTATAA, TATTTATAG, CTATAAATA,
TTATAAAA, TTTTATAA, TTTTATAG, CTATAAAA,
TTATAAAAG, CTTTATAA, TTTTATAG, CTATAAAA,
GGGGGTTGACGTA, TACGTCAACCCCC, TGCGTTAATT₄,
AAAAATTAACGCA,
GGGGGTTGACGTA, TACGTCAACCCCC, TACGTTAATT₄,
AAAAATTAACGTA,

TGACGTATATT₄, AAAAATATACGTCA, GGGGATATGCGTTA,
TAACGCATATCCCC,
TGACGTATATT₄, AAAAATATACGTCA, GGGGATATGCGTTA,
TAACGCATACCCCC,

ATGATTAGTA, TACTAAATCAT, TGTTGAGTTAT, ATAACTCAACA,
GTTAT, ATAAC, ATGAT, ATCAT,

TTACGTGA, TCACGTAA, TTACGTGG, CCACGTAA,
TTACGTGG, CCACGTAA, TTACGTGG, CCACGTAA,
TTACGTGG, CCACGTAA, TTACGTGA, TCACGTAA,
TTACGTGA, TCACGTAA, TTACGTGA, TCACGTAA,
GACGTT, AACGTC, AGCGTT, AACGCT,

TGACGTGT, ACACGTCA, ATACGTTA, TAACGTAT,
TGACGTGG, CCACGTCA, TTACGTTA, TAACGTAA,
CGGTTATTTG, CAAAATAACCG, TAAGATGGTCG oder CGACCATCTTA

enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer besonders bevorzugten Variante des Verfahrens enthalten die zur Amplifikation verwendeten Oligonukleotide außer den oben definierten Konsensussequenzen mehrere Positionen, an denen entweder irgendeine der drei Basen G, A und T oder irgendeine der Basen C, A und T vorhanden sein kann.

In einer besonders bevorzugten Variante des Verfahrens enthalten die zur Amplifikation verwendeten Oligonukleotide außer einer der oben beschriebenen Konsensussequenzen nur maximal zusätzlich so viele weitere Basen, wie es zur gleichzeitigen Amplifikation von mehr als einhundert verschiedenen Fragmenten pro Reaktion aus der chemisch wie oben behandelten DNA erforderlich ist.

In einem dritten Verfahrensschritt wird nun der Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide untersucht.

In einer besonders bevorzugten Variante des Verfahrens erfolgt die Analyse durch Hybridisierung der bereits in der Amplifikation mit einem Fluoreszenzmarker versehenen Fragmente an einen Oligonukleotid- Array (DNA Chip). Der Fluoreszenzmarker kann entweder über die verwendeten Primer oder aber durch ein fluoreszenzmarkiertes Nukleotid (z. B. Cy5-dCTP, kommerziell von Amersham- Pharmacia erhältlich) eingeführt werden.

Dabei hybridisieren komplementäre Fragmente an die jeweiligen auf der Chipoberfläche immobilisierten Oligomere, nicht komplementäre Fragmente werden in einem oder mehreren Waschschriften entfernt. Die Fluoreszenz an den jeweiligen Hybridisierungsorten auf dem Chip erlaubt dann den Rückschluß auf den Sequenzkontext der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.

In einer weiteren bevorzugten Variante des Verfahrens werden die amplifizierten Fragmente auf einer Oberfläche immobilisiert und anschließend eine Hybridisierung mit einer kombinatorischen Bibliothek von unterscheidbaren Oligonukleotid- oder PNA-Oligomer-Sonden durchgeführt. Wiederum werden nicht komplementäre Sonden durch einen oder mehrere Waschschrifte entfernt. Die hybridisierten Sonden werden entweder über ihre Fluoreszenzmarker detektiert oder in einer weiteren besonders bevorzugten Variante des Verfahrens mittels Matrix-assistierter Laser- Desorptions/Ionisations Massenspektrometrie (MALDI-MS) anhand ihrer eindeutigen Masse nachgewiesen. Dabei werden die Sonnenbibliotheken derart synthetisiert, daß die Masse eines jeden Bestandteils eindeutig seiner Sequenz zugeordnet werden kann.

Die Amplifikate können zudem in einer weiteren bevorzugten Variante des Verfahrens hinsichtlich Ihrer durchschnittlichen Größe durch Veränderung der Kettenverlängerungszeiten im Amplifikationsschritt beeinflußt werden. Da hier vorwiegend kleinere Fragmente (ca. 200-500 Basenpaare) untersucht werden, ist eine Verkürzung der Kettenverlängerungsschritte z. B. einer PCR sinnvoll.

In einer weiteren bevorzugten Variante des Verfahrens werden die Amplifikate durch

Gelelektrophorese aufgetrennt, und die Fragmente im gewünschten Größenbereich werden vor Ihrer Analyse ausgeschnitten. In einer weiteren besonders bevorzugten Variante werden die aus dem Gel ausgeschnittenen Amplifikate unter Verwendung des gleichen Satzes an Primern erneut amplifiziert. Dabei können dann nur noch Fragmente der gewünschten Größe entstehen, da Andere als Templat nicht mehr verfügbar sind.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Kit, enthaltend mindestens zwei Primerpaare, Reagenzien und Hilfsstoffe für die Amplifikation und/oder Reagenzien und Hilfsmittel für die chemische Behandlung und/oder eine kombinatorische Sondenbibliothek und/oder einen Oligonukleotid-Array (DNA-Chip), soweit sie für die Durchführung des erfindungsgemäßen Verfahrens erforderlich oder dienlich sind.

Die folgenden Beispiele erläutern die Erfindung.

Beispiele:

Beispiel 1:

Primer zur bevorzugten Amplifikation von CG reichen Regionen im Humangenom

Bei den CG reichen Regionen im Humangenom handelt es sich um sogenannte CpG-islands, die eine regulatorischen Funktion besitzen. Wir definieren CpG Islands derart, dass sie mindestens 500 bp umfassen sowie einen GC-Gehalt von >50% aufweisen , ausserdem ist der Quotient CG/GC > 0,6. Unter diesen Bedingungen liegen 16 Mb als CpG Islands vor. Damit liegen etwa 0,5 % der Genomsequenz in diesen CpG islands, wenn man auch noch jeweils eine Region bis 1000 bp downstream zusätzlich betrachtet. Dieser Überlegung liegen Daten aus der Ensembl Database vom 31.10.00, Quelle Sanger Centre, zugrunde. Die dort verfügbare Sequenz umfasste ca. 3,5 GB, und für die Berechnungen wurden die Repeats maskiert.

Statistisch wäre es bei 12meren zu erwarten, dass sie nur 0,005 mal so häufig an eine der CG-reichen Regionen hybridisieren wie an eine andere beliebige Region im

Genom. Es wurden nun Primer gefunden, welche 1,8 mal häufiger an eine CG reiche Region binden. Zudem ergibt sich mit den entsprechend gefundene Reverse Primer nahezu eine Spezifität für diese CpG islands.

In diesem Beispiel sind die Primer AGTAGTAGTAGT (Seq. ID 1) AAAACAAAAAACC (Seq. ID 2) und alternativ AGTAGTAGTAGT (Seq. ID 19) und ACAAAAAACTAAA (seq. ID 20). Das erste Primerpaar führt mindestens zu den Amplifikaten Seq. ID 3 bis 18, das zweite Primerpaar zu den Amplifikaten der Seq. ID 21 bis 31.

Beispiel 2:

Berechnung der Vorhersage der Anzahl von Amplifikaten in Genomischen Regionen.

Gemäß Anspruch 8 im Patent wird gezeigt mehr als doppelt so viele Amplifikate erstellen zu können, als es statistisch zu erwarten wäre nach Formel 1.

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Formel 1.

F gibt dabei die Anzahl der Vorhergesagten Amplifikate an, die zu erwarten sind, wenn man N Basen als Datenbasis aus dem Genom betrachtet. P ist die jeweilige Wahrscheinlichkeit für die Hybridisierung eines Primeroligonukleotids, getrennt nach Hybridisierung im Sense- und Antisense-Strang. M ist die maximal zulässige Länge der zu erwartenden Amplifikate.

Die Wahrscheinlichkeit P wird bestimmt durch eine Markov Kette erster Ordnung. Dabei wird die Annahme gemacht, dass die DNA eine Zufallsfolge in Abhängigkeit benachbarter Basen ist. Für die Berechnung einer Markovkette sind die Übergangswahrscheinlichkeiten von benachbarten Basen notwendig. Diese wurden empirisch aus 12% des assemblierten humanen Genoms, das vollständig mit Bisulfit behandelt wurde, ermittelt und in Tabelle 1 zusammengefasst. In Tabelle 2 sind die Übergangswahrscheinlichkeiten für den entsprechenden komplementär reversen

Strang angegeben. Diese ergeben sich durch einfaches Vertauschen der Einträge aus der Tabelle 1.

Tabelle 1

Von\nach	A	C	G	T
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

mit

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge) $P_{rbDNA}(von; nach)$

Tabelle 2

Von\nach	A	C	G	T
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A) = 0.4850$$

$$P_{rbDNA}(C) = 0.2199$$

$$P_{rbDNA}(G) = 0.0140$$

$$P_{rbDNA}(T) = 0.2811$$

Damit hängt die Wahrscheinlichkeit, dass sich für einen Primer *PrimE* (mit der

Basenfolge $B_1 B_2 B_3 B_4 \dots$; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots$$

(Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots$$

(anti-sense-Strang zu einem Bisulfit-DNA-Strang);

für einen Primer $Prim$ auf dem sense-Strang ergeben sich

$$N * P_s(Prim)$$

perfekte Basenpaarungen - Werden mehrere Primer ($PrimU, PrimV, PrimW, PrimX$, etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) &= P_s(PrimU) \\ &+ (1 - P_s(PrimU)) P_s(PrimV) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV)) P_s(PrimW) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW)) P_s(PrimX) \\ &+ \dots \end{aligned}$$

(PrimU, PrimV, PrimW... sind hier verschiedene Primer mit unterschiedlichen Basenpaarungen)

und damit als Anzahl der zu erwartenden perfekten Basenpaarungen mit irgendeinem der Primer

$$N * P_s(Primers)$$

Für die Bestimmung von $P_a(Primers)$ auf dem anti-sense-Strang werden die analogen Gleichungen verwendet.

Für das Beispiel mit zwei Primern (einem sense-Primer und einem antisense-Primer) ergeben sich folgende Wahrscheinlichkeiten:

$$P(AGTAGTAGTAGT) = 0.000000860027$$

$$P(AACAAAAACTAA) = 0.000030005828$$

Auf den CpG-Islands, die insgesamt ca. 30.000.000 Basen enthalten, erwartet man eine Häufigkeit von Hybridisierungen für:

AGTAGTAGTAGT: 25.80 auf dem sense Strang

AACAAAAACTAA: 900.17 auf dem komplementär reversen Strang.

Auf den jeweils anderen Strängen können die Primer nicht hybridisieren, da auf dem sense-Strang durch die Bisulfitbehandlung keine Cs außerhalb des Kontextes CG auftreten und entsprechend komplementär auf dem antisense-Strang.

Ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge M ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet, die Wahrscheinlichkeit dafür ist

$$P_a(\text{Primers}) \sum_{i=0}^{M-2} (1 - P_a(\text{Primers}))^i ;$$

für große M und kleine $P_a(\text{Primers})$ wird dieses durch folgenden Ausdruck berechnet:

$$\frac{P_a(\text{Primers})}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] ;$$

für die Gesamtzahl F der Amplifikate, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1] \quad \text{Formel 1}$$

Für das oben angegebene Beispiel ergeben sich für die CpG-Islands mit 30 Mega Basen 3.0498 Amplifikate. Wir können jedoch zeigen (siehe Beispiel 1), dass man mit Primern, die für bestimmte Regionen spezifisch sind, mehr als statistisch vorhergesagte Amplifikate erzeugen kann.

Patentansprüche

1. Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA, dadurch gekennzeichnet, daß man folgende Schritte ausführt:
 - a) in einer genomischen DNA Probe wandelt man durch chemische Behandlung an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um;
 - b) aus dieser chemisch behandelten genomischen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind, gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer, wobei diese Primer jeweils Sequenzen aus an der Genregulation beteiligten und/oder transkribierten und/oder translatierten genomischen Sequenzen enthalten, wie sie nach einer Behandlung gemäß Schritt a) vorliegen würden;
 - c) man bestimmt den Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß man die chemische Behandlung mittels einer Lösung eines Bisulfits, Hydrogensulfits oder Disulfits durchführt.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) verwendeten Oligonukleotide weniger Nukleobasen enthält als es statistisch für eine sequenzspezifische Hybridisierung an die chemisch behandelte genomische DNA Probe erforderlich wäre.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) des Anspruchs 1 verwendeten Oligonukleotide

kürzer als 18 Nukleobasen ist.

5. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) des Anspruchs 1 verwendeten Oligonukleotide kürzer als 15 Nukleobasen ist.
6. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß man in Schritt b) des Anspruchs 1 mehr als 4 verschiedene Oligonukleotide gleichzeitig für die Amplifikation verwendet.
7. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß man in Schritt b) des Anspruchs 1 mehr als 26 verschiedene Oligonukleotide gleichzeitig für die Amplifikation verwendet.
8. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als nach berechnet nach Formel 1 aus an der Regulation von Genen beteiligten Genomabschnitten, wie Promotoren und Enhancern, stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist wie berechnet nach Formel 1,

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Formel 1

wobei man die Berechnung wie folgt durchführt:

bei der mit Bisulfit behandelten DNA kann C nur noch im Kontext CG auftreten, so wird angenommen, daß die primäre DNA eine Zufallsfolge mit Abhängigkeit direkt benachbarter Basen ist (Markov-Kette erster Ordnung); die empirisch aus der Datenbank (vollständig methyliert; mit Bisulfit behandelt) ermittelten

paarweisen Basenwahrscheinlichkeiten ergeben sich gleich für beide DNA-Stränge als $P_{bDNA}(\text{von}; \text{nach})$ aus der folgenden Tabelle:

Von\nach	A	C	G	T
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

mit

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge) $P_{rbDNA}(\text{von}; \text{nach})$

Von\nach	A	C	G	T
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A) = 0.4850$$

$$P_{rbDNA}(C) = 0.2199$$

$$P_{rbDNA}(G) = 0.0140$$

$$P_{rbDNA}(T) = 0.2811$$

; damit hängt die Wahrscheinlichkeit, daß sich für einen Primer *PrimE* (mit der Basenfolge $B_1 B_2 B_3 B_4 \dots$; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots$$

(Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots$$

(anti-sense-Strang zu einem Bisulfit-DNA-Strang);

für einen Primer *Prim* auf dem sense-Strang ergeben sich

$$N * P_s(Prim) ;$$

perfekte Basenpaarungen - Werden mehrere Primer (*PrimU*, *PrimV*, *PrimW*, *PrimX*, etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) &= P_s(PrimU) \\ &+ (1 - P_s(PrimU)) P_s(PrimV) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV)) P_s(PrimW) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW)) P_s(PrimX) \\ &+ \dots \end{aligned}$$

und damit als Anzahl der zu erwartenden perfekten Basenparungen mit irgendeinem der Primer

$$N * P_s(Primers) ;$$

für die Bestimmung von $P_a(Primers)$ auf dem anti-sense-Strang werden die analogen Gleichungen verwendet; ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge M ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet, die Wahrscheinlichkeit dafür ist

$$P_a(Primers) \sum_{i=0}^{M-2} (1 - P_a(Primers))^i ;$$

für große M und kleine $P_a(Primers)$ wird dieses durch folgenden Ausdruck berechnet:

$$\frac{P_a(Primers)}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] ;$$

für die Gesamtzahl F der Amplifikate, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Formel 1

9. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammt, die in mindestens einer Zelle des jeweiligen Organismus in mRNA transkribiert werden, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
10. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus nach der Transkription in mRNA gesplittenen Genomabschnitten (Exons) stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
11. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche für Teile einer oder mehrerer Genfamilien kodieren, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
12. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche für sogenannte „matrix attachment sites“ (MARs)- charakteristische Sequenzen

enthalten, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.

13. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche als sogenannte „boundary elements“ die Verpackungsdichte des Chromatins organisieren, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
14. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus „multiple drug resistance gene“ (MDR)-Promotoren oder kodierenden Regionen stammen, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
15. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß zur Amplifikation der in Anspruch 1 beschriebenen Fragmente zwei Oligonukleotide oder zwei Klassen von Oligonukleotiden verwendet werden, von denen eines oder eine Klasse außer im Kontext CpG oder CpNpG zwar die Base C enthalten kann, nicht aber die Base G und von denen das andere oder die andere Klasse außer im Kontext CpG oder CpNpG zwar die Base G, nicht aber die Base C enthalten kann.
16. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchgeführt wird, von denen eines eine vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen

würde, wenn ein ebenso langes DNA Fragment, an welches einer der Transkriptionsfaktoren

AhR/Arnt	aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator
Arnt	aryl hydrocarbon receptor nuclear translocator
AML-1a	CBFA2; core-binding factor, runt domain, alpha subunit 2 (acute myeloid leukemia 1; aml1 oncogene)
AP-1	activator protein-1 (AP-1); Synonyme: c-Jun
C/EBP	CCAAT/enhancer binding protein
C/EBPalpha	CCAAT/enhancer binding protein (C/EBP), alpha
C/EBPbeta	CCAAT/enhancer binding protein (C/EBP), beta
CDP	CUTL1; cut (<i>Drosophila</i>)-like 1 (CCAAT displacement protein)
CDP	CUTL1; cut (<i>Drosophila</i>)-like 1 (CCAAT displacement protein)
CDP CR1	complement component (3b/4b) receptor 1
CDP CR3	complement component (3b/4b) receptor 3
CHOP-C/EBPalpha	DDIT; DNA-damage-inducible transcript 3/CCAAT/enhancer binding protein (C/EBP), alpha
c-Myc/Max	avian myelocytomatis viral oncogene/MYC-ASSOCIATED FACTOR X
CREB	cAMP responsive element binding protein
CRE-BP1	CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2, CREB2, CREBP1; now ATF2; activating transcription factor 2
CRE-BP1/c-Jun	activator protein-1 (AP-1); Synonyme: c-Jun
CREB	MP responsive element binding protein
E2F	E2F transcription factor (originally identified as a DNA-binding protein essential E1A-dependent activation of the adenovirus E2 promoter)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
Egr-1	early growth response 1
Egr-2	early growth response 2 (Krox-20 (<i>Drosophila</i>) homolog)
ELK-1	ELK1, member of ETS (environmental tobacco smoke) oncogene family
Freac-2	FKHL6; forkhead (<i>Drosophila</i>)-like 6; FORKHEAD-RELATED ACTIVATOR 2; FREAC2
Freac-3	FKHL7; forkhead (<i>Drosophila</i>)-like 7; FORKHEAD-RELATED ACTIVATOR 3; FREAC3
Freac-4	FKHL8; forkhead (<i>Drosophila</i>)-like 8; FORKHEAD-RELATED ACTIVATOR 4; FREAC4
Freac-7	FKHL11; forkhead (<i>Drosophila</i>)-like 9; FORKHEAD-RELATED ACTIVATOR 7; FREAC7

GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-2	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-3	GATA-binding protein 2/Enhancer-Binding Protein GATA2
GATA-X	GATA-binding protein 3/Enhancer-Binding Protein GATA3
HFH-3	FKHL10; forkhead (<i>Drosophila</i>)-like 10; FORKHEAD-RELATED ACTIVATOR 6; FREAC6
HNF-1	TCF1; transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
HNF-4	hepatocyte nuclear factor 4
IRF-1	interferon regulatory factor 1
ISRE	interferon-stimulated response element
Lmo2 complex	LIM domain only 2 (rhombotin-like 1)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
myogenin/NF-1	Myogenin (myogenic factor 4)/Neurofibromin 1; NEUROFIBROMATOSIS, TYPE I
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
NF-E2	NFE2; nuclear factor (erythroid-derived 2), 45kD
NF-kappaB (p50)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p50 subunit
NF-kappaB (p65)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p65 subunit
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NRSF	NEURON RESTRICTIVE SILENCER FACTOR; REST; RE1-silencing transcription factor
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
P300	E1A (adenovirus E1A oncoprotein)-BINDING PROTEIN, 300-KD
P53	tumor protein p53 (Li-Fraumeni syndrome); TP53

Pax-1	paired box gene 1
Pax-3	paired box gene 3 (Waardenburg syndrome 1)
Pax-6	paired box gene 6 (aniridia, keratitis)
Pbx 1b	pre-B-cell leukemia transcription factor
Pbx-1	pre-B-cell leukemia transcription factor 1
RORalpha2	RAR-RELATED ORPHAN RECEPTOR ALPHA; RETINOIC ACID-BINDING RECEPTOR ALPHA
RREB-1	ras responsive element binding protein 1
SP1	simian-virus-40-protein-1
SP1	simian-virus-40-protein-1
SREBP-1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
SRY	sex determining region Y
STAT3	signal transducer and activator of transcription 1, 91kD
Tal-1alpha/E47	T-cell acute lymphocytic leukemia 1/transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TATA	cellular and viral TATA box elements
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
TCF11/MafG	v-maf musculoaponeurotic fibrosarcoma (avian) oncogene family, protein G
TCF11	Transcription Factor 11; TCF11; NFE2L1; nuclear factor (erythroid-derived 2)-like 1
USF	upstream stimulating factor
Whn	winged-helix nude
X-BP-1	X-box binding protein 1 oder
YY1	ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins

bindet, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

17. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß man die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchführt, von denen eine eine vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

18. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchgeführt wird, von denen mindestens eine der Sequenzen (von 5' nach 3')

TCGCGTGTA, TACACGCGA, TGTACGCGA, TCGCGTACA,
TTGCGTGTT, AACACGCAA, GGTACGTAA, TTACGTACC,
TCGCGTGTT, AACACGCGA, GGTACGCGA, TCGCGTACC,
TTGCGTGTA, TACACGCAA, TGTACGTAA, TTACGTACA,
TACGTG, CACGTA, TACGTG, CACGTA,

ATTGCGTGT, ACACGCAAT, GTACGTAAT, ATTACGTAC,
ATTGCGTGA, TCACGCAAT, TTACGTAAT, ATTACGTAA,
ATCGCGTGA, TCACCGCAT, TTACCGCAT, ATCGCGTAA,
ATCGCGTGT, ACACCGCAT, GTACCGCAT, ATCGCGTAC,
TGTGGT, ACCACA, ATTATA, TATAAT,

TGAGTTAG, CTAACTCA, TTGATTAA, TAAATCAA,
TGATTTAG, CTAAATCA, TTGAGTTA, TAACTCAA,

TTTGGT, ACCAAA, ATTAAA, TTTAAT,
TGTGGA, TCCACA, TTTATA, TATAAA,
TTTGGGA, TCCAAA, TTTAAA, TTTAAA,
TGTGGT, ACCACA, ATTATA, TATAAT,

ATTAT, ATAAT, GTAAT, ATTAC,
ATTGT, ACAAT, GTAAT, ATTAC,

GAAAG, CTTTC, TTTTT, AAAAAA,
GTAAT, ATTAC, ATTGT, ACAAT,
GAAAT, ATTTTC, ATTTTT, AAAAT,
GTAAG, CTTAC, TTTGT, ACAAA,
TTAATAATCGAT, ATCGATTATTAA, ATCGATTATTGG, CCAATAATCGAT,
ATCGATTA, TAATCGAT, TAATCGAT, ATCGATTA,

ATCGATCGG, CCGATCGAT, TCGATCGAT, ATCGATCGA,
ATCGATCGT, ACGATCGAT, GCGATCGAT, ATCGATCGC,

TATCGATA, TATCGATA, TATCGGTG, CACCGATA,
TATTAATA, TATTAATA, TATTGGTG, CACCAATA,

GTGTAATATT, AAATATTACAC, GGGTATTGTAT, ATACAATACCC,
GTGTAATTTT, AAAAATTACAC, GGGGATTGTAT, ATACAATCCCC,
ATGTAATTTT, AAAAATTACAT, GGGGATTGTAT, ATACAATCCCC,
ATGTAATATT, AAATATTACAT, GGGTATTGTAT, ATACAATACCC,
ATTACGTGGT, ACCACGTAAT, ATTACGTGGT, ACCACGTAAT,

TGACGTAA, TTACGTCA, TTACGTAA, TAACGTAA,
TGACGTAA, TAACGTCA, TGACGTAA, TAACGTCA,
TTACGTAA, TTACGTAA, TTACGTAA, TTACGTAA,
TGACGTAA, TAACGTCA, TAACGTAA, TAACGTAA,

TGACGT, ACGTCA, GCGTTA, TAACGC,
TGACGT, ACGTCA, ACGTTA, TAACGT,
TTTCGCGT, ACGCGAAA, GCGCGAAA, TTTCGCGC,
TTTGGCGT, ACGCCAAA, GCGTTAAA, TTTAACGC,

TAGGTGTTA, TAACACCTA, TAATATTG, CAAATATTA,
TAGGTGTTT, AAACACCTA, GAATATTG, CAAATATTC,

GTAGGTGG, CCACCTAC, TTATTTGT, ACAAATAA,
GTAGGTGT, ACACCTAC, ATATTTGT, ACAAATAT,

TGC GTGGGCGG, CCGCCCACGCA, TCGTTTACGTA, TACGTAAACGA,
TGC GTGGGCGT, ACGCCCACGCA, ACGTTTACGTA, TACGTAAACGT,

TGC GTAGGC GT, ACGCCTACGCA, ACGTTTACGTA, TACGTAAACGT,
TGC GTAGGC GG, CCGCCTACGCA, TCGTTTACGTA, TACGTAAACGA,
ATAGGAAGT, ACTTCCTAT, ATTTTTGT, ACAAAAAAT,

TCGGAAGT, ACTTCCGA, ATTTTCGG, CCGAAAAT,
TCGGAAGT, ACTTCCGA, GTTTTCGG, CCGAAAAC,
TCGGAAAT, ATTTCCGA, ATTTTCGG, CCGAAAAT,
TCGGAAAT, ATTTCCGA, GTTTTCGG, CCGAAAAC,
GTAAATAA, TTATTTAC, TTGTTTAT, ATAAACAA,
GTAAATAAATA, TATTTATTTAC, TGTTTATTTAT, ATAAATAAACA,

AAAGTAAATA, TATTTACTTT, TGTTTATTTT, AAAATAAACA,
AATGTAAATA, TATTTACATT, TGTTTATATT, AATATAAACA,
TAAGTAAATA, TATTTACTTA, TGTTTATTTA, TAAATAAACA,
TATGTAAATA, TATTTACATA, TGTTTATATA, TATATAAACA,

ATAAATA, TATTTAT, TGTTTAT, ATAAACA,
ATAAATA, TATTTAT, TATTTAT, ATAAATA,
GATA, TATC, TATT, AATA,

TAGATAA, TTATCTA, TTATTTG, CAAATAA,
TTGATAA, TTATCAA, TTATTAG, CTAATAA,
GATAA, TTATC, TTATT, AATAA,

GATG, CATC, TATT, AATA,

GATAG, CTATC, TTATT, AATAA,
GATAAG, CTTATC, TTTATT, AATAAA,

TGTTTATTAA, TAAATAAAACA, TAAATAAAATA, TATTTATTAA,
TGTTTGTAA, TAAACAAACA, TAAATAAAATA, TATTTATTAA,
TATTTATTAA, TAAATAAAATA, TAAATAAAATA, TATTTATTAA,
TATTTGTAA, TAAACAAATA, TAAATAAAATA, TATTTATTAA,

GTTAATGATT, AATCATTAAAC, AATTATTAAT, ATTAATAATT,
GTTAATTATT, AATAATTAAC, AATAATTAAT, ATTAATTATT,
GTTAATTAAT, ATTAATTAAC, ATTAATTAAT, ATTAATTAAT,
GTTAATGAAT, ATTCACTAAAC, ATTTATTAAT, ATTAATAAAAT,

TAAAGTTA, TAAACTTTA, TGAAATTTG, CAAAATTCA,
TAAAGGTTA, TAACCTTTA, TGATTTTG, CAAAAATCA,

AAAGTGAAATT, AATTCACCTT, GGTTTTATTTT, AAAATAAAACC,
AAAGCGAAATT, AATTCGCTTT, GGTTTCGTTT, AAAACGAAACC,

TAGTTTATTTTTT, AAAAAAAATAAAACTA, GGGAAAGTGAAATTG,
CAATTCACCTTCCC,
TAGTTTATTTTTT, AAAAAAAATAAAACTA, GGAAAAGTGAAATTG,
CAATTCACCTTCC,
TAGTTTTTTTTTT, AAAAAAAAAAAACTA, GGAAAAGAGAAATTG,
CAATTCTCTTTCC,
TAGTTTTTTTTTT, AAAAAAAAAAAACTA, GGGAAAGAGAAATTG,
CAATTCTCTTTCCC,
TAGGTG, CACCTA, TATTG, CAAATA,

TTTAAAAATAATTT, AAAATTATTTTAAAA, AGGGTTATTTTAGAG,
CTCTAAAAATAACCCCT,
TTTAAAAATAATTT, AAAATTATTTTAAAA, GGAGTTATTTTAGAG,
CTCTAAAAATAACTCC,
TTTAAAAATAATTT, AAAATTATTTTAAAA, AGAGTTATTTTAGAG,
CTCTAAAAATAACTCT,
TTTAAAAATAATTT, AAAATTATTTTAAAA, GGGGTTATTTTAGAG,
CTCTAAAAATAACCCC,

TGTTATTAAAAATAGAAA, TTTCTATTTTAATAACA,
TTTTTATTTTAGTAATA, TATTACTAAAAATAAAAAA,
TGTTATTAAAAATAGAAT, ATTCTATTTTAATAACA,
GTTTATTTTAGTAATA, TATTACTAAAAATAAAAC,
TTGGTAT, ATACCAAA, GTGTTAAA, TTTAACAC
GGGA, TCCCC, TTTT, AAAAAA,

TAGGGG, CCCCTA, TTTTTA, TAAAAA,
GAGGGG, CCCCTC, TTTTTT, AAAAAA,
TGTTGAGTTAT, ATAACCTCAACA, ATGATTTAGTA, TACTAAATCAT,
TGTTGATTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,
TGTTGAGTTAT, ATAACCTCAACA, ATGATTTAGTA, TACTAAATCAT,
TGTTGATTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,

GGGGATTTTT, AAAAATCCCC, GGGAAATTTTT, AAAAATTCCC,
GGGGATTTTT, AAAAATCCCC, GGGGATTTTT, AAAAATCCCC,
GGGGATTTTT, AAAAATCCCC, GGAAATTTTT, AAAAATTCC,
GGGAATTTTT, AAAAATTCCC, GGAAATTTTT, AAAAATTCC,
GGGAATTTTT, AAAAATTCCC, GGAAATTTTT, AAAAATTCC,
GGGATTTTT, AAAAAATCCC, GGAAAGTTTT, AAAACTTCC,
GGGAATTTTT, AAAAATTCCC, GGGAAATTTTT, AAAAATTCCC,
GGGATTTTT, AAAAAATCCC, GGGAAAGTTTT, AAAACTTCCC,
GGGATTTTTA, TAAAAAAATCCC, TGGAAAGTTTT, AAAACTTCCA,
TTTAGTATTACGGATAGAGGT, ACCTCTATCCGTAATACTAAA,
GTTTTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAAAAC,
TTTAGTATTACGGATAGAGTT, AACTCTATCCGTAATACTAAA,
GGTTTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAAACC,
TTTAGTATTACGGATAGCGTT, AACGCTATCCGTAATACTAAA,
GGCGTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAACGCC,
TTTAGTATTACGGATAGCGGT, ACCGCTATCCGTAATACTAAA,
GTCGTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAACGAC,

ATATGTAAAT, ATTTACATAT, ATTTGTATAT, ATATACAAAT,
TTATGTAAAT, ATTTACATAA, ATTTGTATAA, TTATACAAAT,

GAATATTTA, TAAATATTC, TGAATATTT, AAATATTCA,
GAATATGTA, TACATATTC, TGTATATTT, AAATATACA,

ATAAT, ATTAT, ATTAT, ATAAT,
GTAAT, ATTAC, ATTAT, ATAAT,

AATGTAAAT, ATTTACATT, ATTTGTATT, AATACAAAT,

ATTTGTATATT, AATATACAAAT, GGTATGTAAAT, ATTTACATACC,
ATTTGTATATT, AATATACAAAT, AATATGTAAAT, ATTTACATATT,
ATTTGTATATT, AATATACAAAT, AGTATGTAAAT, ATTTACATACT,
ATTTGTATATT, AATATACAAAT, GATATGTAAAT, ATTTACATATC,

AGGAGT, ACTCCT, ATTTTT, AAAAAT,
GGGAGT, ACTCCC, ATTTTT, AAAAAT,
GGATATGTTCGGGTATGTTT, AAACATACCCGAACATATCC,
GGATATGTTCGGGTATGTTT, AAACATACCCGAACATATCC,
GGATATGTTCGGGTATGTTT, AAACATACCCGAACATATCC,
AGATATGTTCGGGTATGTTT, AAACATACCCGAACATATCT,
TCGTTTCGTTTAGATAT, ATATCTAAACGAAACGA,
ATATTTAGAGCGGAACGG, CCGTTCCGCTCTAAATAT,

CGTTACGGTT, AACCGTAACG, AATCGTGACG, CGTCACGATT,
CGTTACGGTT, AACCGTAACG, GATCGTGACG, CGTCACGATC,
CGTTACGTT, AAACGTAACG, AAGCGTGACG, CGTCACGCTT,
CGTTACGTT, AAACGTAACG, GAGCGTGACG, CGTCACGCTC,

TTTACGTATGA, TCATACGTAAA, TTATGCGTGAA, TTCACGCATAA,
TTTACGTTGA, TCAAACGTAAA, TTAAGCGTGAA, TTCACGCTTAA,
TTTACGTTTA, TAAAACGTAAA, TGAAGCGTGAA, TTCACGCTTCA,
TTTACGTATT, TAATACGTAAA, TGATGCGTGAA, TTCACGCATCA,
AATTAATTAA, TTAATTAATT, TTGATTGATT, AATCAATCAA,
TATTAATTAA, TTAATTAATA, TTGATTGATG, CATCAATCAA,
TAATTAT, ATAATTA, ATGATTG, CAATCAT,
TAGGTTA, TAACCTA, TGATTTA, TAAATCA,
TTTTAAATATTTT, AAAAATATTTAAAA, GGGGGTGTGTTGGGG,
CCCCAAACACCCCCC,
TTTTAAATTATTTT, AAAATAATTTAAAA, GGGGTGGTTGGGG,
CCCCAAACCACCCC,
TTTTAAATTTTTT, AAAAAAAATTTAAAA, GGGGGGGTTGGGG,
CCCCAAACCCCCC,
TTTTAAATAATTTT, AAAATTATTTAAAA, GGGGTTGTGTTGGGG,
CCCCAAACACAACCC,
GAGGCGGGG, CCCCGCCTC, TTTCGTTTT, AAAACGAAA,
GAGGTAGGG, CCCTACCTC, TTTTGTGTTT, AAAACAAAA,
AAGGCGGGG, CCCCGCCTT, TTTCGTTTT, AAAACGAAA,
AAGGTAGGG, CCCTACCTT, TTTTGTGTTT, AAAACAAAA,
GGGGGCGGGGT, ACCCCGCC, ATTCGTTTT, AAAAACGAAAT,
GGGGGCGGGGT, ACCCCGCC, GTTCGTTTT, AAAAACGAAAC,
TATTATTTAT, ATAAAATAATA, GTGGGGTGATA, TATCACCCCCAC,
GATTATTTAT, ATAAAATAATC, GTGGGGTGATT, AATCACCCCCAC,
ATTACGTGAT, ATCACGTAAT, ATTACGTGAT, ATCACGTAAT,
ATTACGTGAT, ATCACGTAAT, GTTACGTGAT, ATCACGTAAC,
TTTATATGG, CCATATAAAA, TTATATAAGG, CCTTATATAA,
TTATATATGG, CCATATATAA, TTATATATGG, CCATATATAA,
AAATAAT, ATTATTT, GTTGTGTT, AAACAAAC,
AAATTAA, TTAATTT, TTAGTTT, AAACCTAA,
AAATTAT, ATAATTT, GTAGTTT, AAACCTAC,
AAATAAA, TTTATTT, TTTGTGTT, AAACAAA,
ATTTCGGAAATG, CATTCCGAAAAAT, TATTTTCGGGAAAT,
ATTCCCGAAAATA,
ATTTCGGAAATG, CATTCCGAAAAAT, TATTTTCGGGAAAT,
ATTCCCGAAAATA,
ATTTCGGGAAATG, CATTCCCGAAAAT, TATTTTCGGGAAAT,
ATTCCGAAAATA,
ATTTCGGGAAGTG, CACTTCCCGAAAAT, TATTTTCGGGAAAT,

ATTTCCGAAAAATA,

AATAGATGTT, AACATCTATT, AATATTGTT, AACAAATATT,
AATAGATGGT, ACCATCTATT, ATTATTGTT, AACAAATAAT,

GTATAAATA, TATTTATAC, TATTTATAT, ATATAAATA,
GTATAAATG, CATTATAC, TATTTATAT, ATATAAATA,
GTATAAAAA, TTTTTATAC, TTTTTATAT, ATATAAAAA,
GTATAAAAG, CTTTTATAC, TTTTTATAT, ATATAAAAA,
TTATAAATA, TATTTATAA, TATTTATAG, CTATAAATA,
TTATAAATG, CATTATAA, TATTTATAG, CTATAAATA,
TTATAAAAA, TTTTTATAA, TTTTTATAG, CTATAAAAA,
TTATAAAAG, CTTTTATAA, TTTTTATAG, CTATAAAAA,
GGGGGTTGACGTA, TACGTCAACCCCC, TGCGTTAATTTT,
AAAAATTAACGCA,
GGGGGTTGACGTA, TACGTCAACCCCC, TACGTAAATTTT,
AAAAATTAACGTA,

TGACGTATATTTT, AAAAATATACGTCA, GGGGATATGCGTTA,
TAACGCATATCCCC,
TGACGTATATTTT, AAAAATATACGTCA, GGGGGTATGCGTTA,
TAACGCATACCCCC,
ATGATTTAGTA, TACTAAATCAT, TGTTGAGTTAT, ATAACCAACA,
GTTAT, ATAAC, ATGAT, ATCAT,

TTACGTGA, TCACGTAA, TTACGTGG, CCACGTAA,
TTACGTGG, CCACGTAA, TTACGTGG, CCACGTAA,
TTACGTGG, CCACGTAA, TTACGTGA, TCACGTAA,
TTACGTGA, TCACGTAA, TTACGTGA, TCACGTAA,
GACGTT, AACGTC, AGCGTT, AACGCT,

TGACGTGT, ACACGTCA, ATACGTTA, TAACGTAT,
TGACGTGG, CCACGTCA, TTACGTTA, TAACGTAA,
CGGTTATTTG, CAAAATAACCG, TAAGATGGTCG oder CGACCATCTTA

enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

19. Verfahren nach einem der Ansprüche 16 bis 18, dadurch gekennzeichnet, daß die zur Amplifikation verwendeten Oligonukleotide außer den in den Ansprüchen 16 bis 18 definierten Konsensussequenzen mehrere Positionen enthalten, an denen

entweder irgendeine der drei Basen G, A und T oder irgendeine der Basen C, A und T vorhanden sein kann.

- 20.Verfahren nach Anspruch 19, dadurch gekennzeichnet, daß die zur Amplifikation verwendeten Oligonukleotide außer einer der in den Anspruch 18 beschriebenen Konsensussequenzen nur maximal zusätzlich so viele weitere Basen enthalten, wie es zur gleichzeitigen Amplifikation von mehr als einhundert verschiedenen Fragmenten pro Reaktion aus der chemisch behandelten DNA, berechnet entsprechend Anspruch 8, erforderlich ist.
- 21.Verfahren nach einem der vorangehenden Ansprüche dadurch gekennzeichnet, daß die Untersuchung des Sequenzkontextes aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide gemäß Anspruch 1 c) durch Hybridisierung der bereits in der Amplifikation mit einem Fluoreszenzmarker versehenen Fragmente an einen Oligonukleotid- Array (DNA Chip) erfolgt.
- 22.Verfahren nach einem der Ansprüche 1 bis 20 dadurch gekennzeichnet, daß die amplifizierten Fragmente auf einer Oberfläche immobilisiert und anschließend eine Hybridisierung mit einer kombinatorischen Bibliothek von unterschiedbaren Oligonukleotid- oder PNA-Oligomer-Sonden durchgeführt wird.
- 23.Verfahren nach Anspruch 22 dadurch gekennzeichnet, daß die Sonden mittels Matrix-assistierter Laser-Desorptions/Ionisations Massenspektrometrie (MALDI-MS) anhand ihrer eindeutigen Masse nachgewiesen werden und damit der Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide entschlüsselt wird.
- 24.Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die Amplifikation wie beschrieben in Schritt b) des Anspruchs 1 durch eine Polymerase Kettenreaktion ausgeführt wird, in der die Größe der amplifizierten Fragmente mittels auf weniger als 30 s verkürzter Kettenverlängerungsschritte

begrenzt wird.

25.Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß nach der Amplifikation gemäß Schritt b) des Anspruchs 1 die Produkte durch Gelelektrophorese aufgetrennt werden und die Fragmente, die kleiner als 2000 Basenpaare oder kleiner als ein beliebiger Grenzwert unterhalb von 2000 Basenpaaren sind, durch Ausschneiden von den anderen Produkten der Amplifikation vor der Auswertung gemäß Schritt c) des Anspruchs 1 abgetrennt werden.

26.Verfahren nach Anspruch 25 dadurch gekennzeichnet, daß nach der Abtrennung von Amplifikaten bestimmter Größe diese vor der Durchführung des Schrittes c) des Anspruchs 1 nochmals amplifiziert werden.

27.Kit, enthaltend mindestens zwei Primerpaare, Reagenzien und Hilfsstoffe für die Amplifikation und/oder Reagenzien und Hilfsmittel für die chemische Behandlung gemäß Anspruch 1 a) und/oder eine kombinatorische Sondenbibliothek und/oder einen Oligonukleotid-Array (DNA-Chip), soweit sie für die Durchführung des erfindungsgemäßen Verfahrens erforderlich oder dienlich sind.

SEQUENZPROTOKOLL

ALLGEMEINE ANGABEN:

ANMELDER:

NAME: Epigenomics AG
STRASSE: Kastanienallee 24
LAND: Berlin
POSTLEITZAHL: 10435
TELEFON: 030-243450
TELEFAX: 030-24345555

BEZEICHNUNG DER ERFINDUNG: Verfahren zur parallelen Detektion
des Methylierungszustandes von
genomischer DNA

ANZAHL DER SEQUENZEN: 31

COMPUTERLESBARE FASSUNG:

DATENTRÄGER: Diskette
COMPUTER: IBM PC-kompatibel
BETRIEBSSYSTEM: PC-DOS/MS-DOS

DATEN DER JETZIGEN ANMELDUNG:

ANMELDENUMMER: nicht bekannt
ANMELDETAG: 6.12.2000

ANGABEN ZU SEQ ID-NO:1:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:1:

AGTAGTAGTA GT

12

ANGABEN ZU SEQ ID-NO:2:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:2:

AAAACAAAAAA CC

12

ANGABEN ZU SEQ ID-NO:3:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 973 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:3:

AGTAGTAGTA GTAGCGTTT TGAAGTTTT TGTGGAAGGT GAGAAATT A TCGATAAGTT	60
TTTTAGTAGG AGTGTGTTG GGGAGGGAGT GAGTGGGAGA TTAGAAATGG GGTGCGGTGGA	120
ATATTTTG T AAAATTAG GAATTATTAG TATTTTATT TTTTTATAA AGATTGTTT	180
TATGTTGAG TTGTTTATA TAGTGTAGAA ATTGAGAATG TAAACGTGTA AACGATCGGT	240
GTGATTATT GAAAGGCAT GCGCGTTTT TTTTGTAGG TAGAATTGTT TTAGGAAGTA	300
TGAAAGTCCA CGTAGTCGG TGTTAGGTGT TAGCGATTG AGTTCGTTT CGGGTTTAT	360
TTATTTTAG TTCGGTTTT AGATATTTT CGAGGCCTT TTTTTGTTC GTTCGATT	420
TGAGCGGAGC GTTCGGGGT GTGAGGAGAA TCGGTAAATT TTGCGGGCG TTGGCGCTCG	480
TAGAGTCGTC GCGCGTTAG TTTCGTTACG TTGGTTGTCG AGCGTATTG GGTGCGTT	540
GGCGGGGAGG CGTCGAGGGT AGTTAAGGGG AGTAGGTTAC GTGAGGAGAG GAGTTTGATT	600
TATTTTTAG GCGGTAGGCG TATGCGTATT TTTTATTG TGTTCGGTGC GGAGGTTTAC	660
GTGGAATCGT ACGTGTGTTGG TTTGTAGTT AGGGGTTTT GGTCGGGGC GCGTAGGGC	720
GGGTCGTAG TGGGATTGCG GGAGAGGGC GCGCGGGGC GGAGCGTTG GAGATTTAGT	780
TTGCGGGTT CGTAATTATT ATTGCGATA ATTATAGTT TTGGAGAGTT GTTGTAGTT	840
TTTGCAGGGT TTTTACGAA TTTATTATAT TTAATATTG TGAATTAGA AATTAAGATA	900
AAGATTGTAT TTTGTGTTG TAAATTATTA GTTCGTAAT TTAGAATT A TTTGTAAAT	960
GGGTTTTGT TTT	973

ANGABEN ZU SEQ ID-NO:4:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1890 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:4:

AGTAGTAGTA GTTAAGTTA AGAGTGATAT TTTTTAGAT ATAGTATAAT ATTGTTTTG	60
TAAAAGGTTA TTTATATGAA TATAATGTT TTTGGGTTAA TTAATTGTT TTGAGAATAG	120
TTAGGTTAAA ATTTTAGGT TTTTATT TTT GGGTTATTTA AGTTAGGATC GCGACGTGAG	180
TTCGGGGTGA GTCGAGGGGT ATTTCGGGTC GAGGTTATT C GAATCGAAGT TATGAGGCGC	240
GGGTCGTAA TCGGAAGTGG TTTAGGGAGA GTTGTACGAG ATTGCGGGT TGTGATTG	300
AAATAAAATA AAATAAAATA AAATTTAAT TGTTTGGG TTGATTTTA AAAGAGCGTT	360
TTTTGGTTT AAGAGGTCGG CGTCGCGGA GATGCGTT TAAAGGGTTG TTTTTAACG	420
GTGTAGGTCG CGTACGGGT TTTTAGCGG CGGGTAAAA TGGGCGTCGG TATTGCGGAG	480
GCGTTGTTT AGGCGTCGG GCGTCGTTA TAGAGTACGT TCGTTGCGG TTTAGAGCG	540
TCGTTTTT GTCGTTTCG TCGTTCGGT TGACGTTCGG ATCGCGTCG GTTATCGTT	600
TTCGTTTCGA CGGTTACGTT TGTTTAAAT CGCGCGGGCG TTTTTAGGT GTGTTGGC	660
GGGTTTCGTT TCGTGTGTTT AAGGTTCGTT TCGCGCGTT AGTCGCGCGT TCGTTGTTT	720
TTTTCGTT TTATAGTTT GTTTTATAG TTCGTTCGT TTTTAAGTT TCGTTTTTA	780
GGAATTGCG CGTCGAAGGT TAGGTTGGG CGGAGCGTAT AGCGTTGGC GTGGGGAGG	840
TTGCGTCGTA GTATTCGGTT GGTTAGGATT AAGTGGGTT GAGGCGGACG TGAGAAGGGT	900
CGGGTTAAGA TGGCGGTGTA GGTGGTAG CGGGTGTAG CGGTTATT CGAGTTGAC	960
GTTTTTTCG TTTGTTTAA TTACGTTTG AGTATAGAGA AGGAGGAAGT AATGGGTTG	1020
TGTATAGGGG AGGTGAGTAG GTTGTAGT TTGGATGAA TTTGTTGAG TAGTTTAGT	1080
GTGTTTCGG GGTGGGTGTC GGTAGTTTT AGGGTTGCGG AGGTTATAGG TATTTTCGAT	1140
TTAGGTTTT GGATATTTT TATAGTGT TGTGCGGGG AGTTCGTT TTTTGGTCG	1200
TTTGATATT TTTAGTTT TTGCGAGTT TCGCGGGTT GTATAGGTT CGGGAGTTT	1260
TTTTTTTT TTGTTAGTTG GGGTTTTT TTAGATTG TGTTCGTT CGGTTGATTA	1320
AAATTGTAAG GTAGGTTAGA AAGATATTGG AGTATAATGA GGATGTTCGG GTATTACGAC	1380
GTAGTAGTTG ATTATAGTTA GAGTTTTG TTTCGTTCG AGTTTTGTT TTAGGGAGTA	1440
GAGATATTAA TTAAAGTATT TGAAGGGTAT CGAAGAGTTT AATAGAAGGT GCGGGGTTG	1500
AAGGAAGTAA AAGTTTCGT TGTATTGTGT TGAGGAGGGG TCGAAGAGGA TGAGGAAATA	1560
TAGTTTAGTT GTTATAGTT TAAAGTAATT TTTAGTTTT TTATATTATG TGCCTGAATA	1620
TATGATTAA TTGTTATATA ATTTGTATT ATATATGTTA AATAAACGTA ATGTGATTAA	1680

TAATGTTTT TGTTCGGAGA GTAAAGTGGGT TTTATGAAA GTGTTATTG TTATATATTG	1740
TTTATTATA TTTTATAATT TTTAAGTTAT TGTTAATTAA AAATTATGGC GATTTGGTA	1800
GTATTGTGTT CGGATCGTGT TTTAAATTAA TTATTATTAT TATTCGGGG TTTTTTAAT	1860
GATATAATTG TGAATTTGG TTTTGTTTT	1890

ANGABEN ZU SEQ ID-NO:5:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2222 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:5:

AGTAGTAGTA GTTTGGAGA GTAAAGTGGGT TTTATGAAA GTGTTATTG TTATATATTG	60
TTATTTTAN AGGATTTAG AATATTTCG AATTCGAAAA TAGAGTGAGT GTGGAGGGGA	120
GGGGGGTGTT TACGTGGAGG AGGGTCGCAGA GAAGGAGAGT GCGTCGGGTT GATGGTTAGC	180
GGTGTTCGG GGGTCGATGT CGGGCGGGAT CGGCGGCCGCG GGGGTGGGGC GACGGCGGGG	240
CGGCCTGCGA GTAGTAGAGG GGCCTTCGT AGAGGTCGGG GGGGGCGCGG GTTTCGCGT	300
CGTTTGTGTA GTAGTCGTT CGTCGTTTT TTCGGTTCGT TCGTAATCGT CGCAGGATTG	360
GTCGGTCGG GTTTTGAA TTTTACGG GGGATAGTCG ATCGGGGGCG TCAGGGTTCG	420
TTTCGTATT CGATTCGGAC GGCGTAGTA GGGGGAGGGG TGTTGTATCG AGGAAGTTAA	480
GGTTTATTA TTTTGTTCGG CGGGTCGGCG GTTCGTTGGT TCGTTGGTT GGAGAAAGTGT	540
TGCGTTAGG TTGGTTCGTA GGAAACGGCG GCGGCAGGTT ANTTNTANTT TTNNNNNTNT	600
TNNTTGNAN TNNTTTTTA TAGGGGAGGG GTAGGCAGGTT CGCAGGTTTC GCGTCGGTCG	660
GGCGATTGGG TACCGCAGGG AGCGGGTAGG GTAGGGGGAA ATAAATTAAG GTCGAGATTG	720
AGAGTCGGAG AGCGCGGGAG GAGTAGCGGC GAGAGGTAGG AGAGGTAGAG AAGAAGAAAG	780
GAGGGAGAGA GGGGGCGAAG ACGGTCGGGA AATTCGGGGT TGTAGTTCG CGTCGCCTCG	840
GAGTCGTGAC GGATTATTG TAGCGCTCGC GTTGATTGTT AATTTTAAGG TACGAAAAAA	900
GGGGAGGGTA GAGAGGGAGG GGAAAGCGGA GTGTTGTAGC GTCGGGGCGG GGGCGGGTTT	960
TCGCGAGCGT CGTATATTG GGAATTGTN GTTCGTTGC GGGCGAGTGT CGCGTGTGTT	1020
GGCGAGTTT GTTGGGGAA NTTGAGCGC GCGGATTACG TTCGTTTTT AGTCGGTCGT	1080
TTTTGCGTG TAGGGTGGGG GGAGGTTAGG GCGGGGGCG CCGACGTCCG TGACGTGCG	1140
TGGCGGAGTT TTTCGGTTA TGTGGTTGGA GCGGGCGGG GAAGAAGGTA AGGTTGGAG	1200
GGGAAGGAGG AAATGCGAGG GTTTTCGGC GGGAGGAAGC GCGAGGGGTG CGCGAGTGT	1260
AGCGAAAGGG GGATAGGGGT TTTGGAGTGG AAGTTTGCG GGAGAGGAGG AGTAGAGGAG	1320
TAGTTTAGG GTGTTAGAAT TATTCGGATG TCGTATTAAA AAATAGGAAA AATTTAAGTT	1380
TTCGAGAGG TAGTAAAGG TATTTGTGTA TTCGTGTATC GATTGGATT TATTATAGAA	1440
TTGTACGGTT TATTAGGATT GTTGTATTG CCGTGTAGTA TTCGTAGGTT TATTGTAAT	1500
ATGAATTGGT AGAGTTAGGA ATTAGGTTGT AAAGATAGAG AAGCGAGTTA AAGTTGGTCG	1560
TAGTCGAGG CGTGGGGAGA ATTGGGTAAC CGAGGAGAA GGGATATTG TTATTGCG	1620
AAAGATTTTT TATTTAGTT TGTATTTA TAAATCGTT AGATTTTTT TTGGCGGAGT	1680
TTATTAGTT TTGTTAAAA AAAGAAAAAA ATTCAATGT AGATTGTTCG TTTATTTTT	1740
TAGTAGAGTT ATATTTATTG TTGTGGTAAT TATTTTTA GAAAATTAG ATTATAATAG	1800
GAAATTATAT TTAGAAAAGT ATAAGGAGGA AATTTATGTT CGAGAGAAGA AATAAAAGTTG	1860
TTAGGAGAGG TGTGATGAGG ATAACGAAGA AAATATTG TAATTATTG AATTAGTTAT	1920
TTTTTATAA GTTTGATAA TCGTCGTAT TCGTGTGTGT TTGAGAGTGT GTGTGTGT	1980
GTGTGTGTGT GTGAGAGAGA GAGAGAGGAT GTTTTGAG TGATCGGAAA TTTTTATTG	2040
ATAAAGTTT TGAGTTATT TATAAAAAGT TATTGTGTTG TTTTTTTA AAAAAGTTT	2100
AATGATTTTT ATATGAGAAT TTATTTTT TTATGTGTTA ATTATTTTT TATATGATGA	2160
GTATGTATTG TTTTTAATA AACGTTAGC GATTGTTTT AGGATTAAAT GGTTTTGTT	2220
TT	2222

ANGABEN ZU SEQ ID-NO:6:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 307 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:6:

AGTAGTAGTA	GTTTCGTTAC	GAATACGATT	AGTTTATATT	GGTTGTTAG	TATTCGGTT	60
AGTTTTGTG	GGACGTTGGC	GCGTATATCG	TTTCGTCGGG	GGGTCGGGGA	TTTCGGGGAT	120
TTCGGGGTCG	AGGACGAGGG	AGGCAGTAG	GTCGTTGGTT	TCGACGATT	TGCGATCGAG	180
GCGTTGTAG	TTGTTCGGGG	CGAGGCCGGC	GATTTAGGT	TTTCGAGTAT	TTTCGAGATA	240
TTCGCGTTAA	GTTTGGTTTC	GGGGGTGGTT	GTCGTCGGGT	TTCGTTGCG	TTTTGGTTT	300
					TTGTTT	307

ANGABEN ZU SEQ ID-NO:7:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 523 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:7:

AGTAGTAGTA	GTTGTTGTTG	TTCGAGGTTT	CGGTGTAGTT	GGAAGTTTG	GAGGTGGGGA	60
GAGGGATGTT	AGGTAAGTGG	TACGGCGAGC	GTAAGGGAAG	GGGTTAGTTA	TTGATTAGCG	120
GTAGTAATTG	TAGGAATCGT	CGTCGTAGTT	GTAGTCGTTT	TTTCGTCGT	TTTTTCGGGT	180
TTTCGGGAAA	ATGGTTGTGG	GGTTGGTCGC	GTCGTTAAGT	TTGTTTCGC	GCGGTGAAGA	240
GCGGGTTGTT	TGGGGGAGTC	GTTCGTTAAC	TTTCGCGCGC	GTTCGTTAAC	GCAGGAGTCG	300
TGGTAGGATT	CGAGGGGTTA	CGAGTTGATA	TTTTTTGGG	TTGTATAAAA	AGTTGAGGCG	360
GGCGTTGGGA	GGAGGTAGCG	GTTGTTGCGG	TGCGGTTTTT	TTTTTTTTT	ATATTTGTT	420
CGGGTTATTT	TTTTTTTTT	TTTTTTTCGT	TTTCGTTTT	TTTTTTTACG	CGGGTTTTTC	480
GGGGTTTTGG	CGGTTTCGG	TTGAAGGTCG	CGGTTTTGT	TTT		523

ANGABEN ZU SEQ ID-NO:8:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 653 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:8:

AGTAGTAGTA	GTAGCGTCGG	GAGTTCGCGT	AGAAACGATT	TGATTTTCG	CGGCGTTATT	60
TTTTTTGCG	ATTGGCGTTG	CGCGGAGAAA	ATTTAGTTG	TCGATGTTTC	GTTCATTATT	120
TCGTTTTAT	TTATCGCGCG	GTTCGGGG	AATTGTAGTT	TTTCGTCGAG	CGGTCGGTAG	180
CGCGGAGGTC	GTCGGCGGAT	TTAGCGTTT	AGGAAGTTT	CGCTAGGCCT	AAGCGTTTT	240
TAGGTAATTA	CGTCGCGGTT	CGAGGTTGTA	GATTTGGTA	GGGGATATAA	TTGGAGAATG	300
AGGGTGGTCG	GGTGGGGAGTA	GGGATTAGGT	TTTATTAGA	GTTCGGAGTT	TGGAGTCGT	360
ACGATTCGAG	GTTCGGGGTT	TTAGGTTGTT	TATAGAAAGA	GCGGAAAGTT	TAAGAATCGG	420
GTCGCGTTG	GTTGAGTTG	ATAATGTTTC	GGTTTACGC	GTATTGACGT	CGATTTTGT	480
ATGTTTAGTG	TCGTTGCAGGG	GTTGGTATTG	CGGTTCGGGT	TTGGCGTTGA	GGAGTTGGT	540
TTAGTTCGTT	TTTGGTTTTC	CGGGAGCGTT	TGGTCGTCGG	TAAGTTAGG	TTAGGTTGTT	600
TCGGGGACGT	AGGGTGCAGGT	AGACGTTTTT	TTACGTTTCG	GGGTTTTGT	TTT	653

ANGABEN ZU SEQ ID-NO:9:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1461 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:9:

AGTAGTAGTA	GTAGGGCGGTG	GTTAGGTCGG	GTCGGGCGGT	TATTAGCGGT	TTTCGGAGGT	60
ATTTAGGTTT	TACGGTCGAG	TTTTTGGTCG	GAGATCGGTC	GTTTACGGGG	GGTTATAGTA	120
GCGGTCTTC	GTAGGATG	GAGAGGCAGG	TTTTAGGTTG	GGTCGGAGC	GAGTTTTTA	180
GGGTTTGTG	ATACGGCGG	GTTCGGTGGT	CGGTATTG	TTCGTACGTG	TCGAGGGGT	240
TTTCGGGTTT	TCGGTATTAT	GGTTATTATC	GGGGTTTCGA	TTACGACGAG	GTCGATGGTT	300
CGGGTAGCGG	GGGCAGGCGAG	GAGGTTATGG	TCGGGGTTA	CGACGCGTTA	TTTTCGTAC	360
GATAACCGTT	TTCGGGCGTT	ATCGGGCGTT	CGTTAGGAT	TTTCGGGTT	TCGGGTTCGG	420
TTTGCCTTTC	GTTCGGGCGG	TACGGTCGGC	GATTTTTAA	CGGTTATTAT	TCGGCGTACG	480
GATTGGTTAG	GTTCAGGCGG	TCGGGGTTA	GGAAAGGGTT	GTACGAATT	TATAGCGAGA	540
GTGACGATGA	TTGGTGTAA	GTTCAGGCGA	GGTGGCGTT	GTTCGGTTT	TTACGTATT	600
TACGTATATA	TTTATTTCGA	GGAGTCGCGT	AGAGGTCGCG	GGGGTTTAGT	ATAGAGGGTT	660
CGGGAGAGGG	TTAGTCGGG	GATTTTAGAT	TTTGGAGAGG	TTAGGGTTGG	GTTATAAGGG	720
TGTTTCGTAG	AGATTTCCG	TTAAAAGAGA	TTTTTTGGG	AGTTACGGC	GTTCGGTTAAT	780
TAGTTTCGAT	TTTTTATTT	ACGATAGGGG	TTTCGGGTG	GGAGGTAGGG	AGTAGATAAA	840
TTATATAGTT	AAGGGATTG	AATTAATT	GTTATTTTG	GAGAATT	GGGAATATGA	900
AAAAAAAAAA	AAAAAA	AAAAAAATA	TTTTAAAG	AAAAAACGGG	GAGAAAAAAA	960
TAGTTTTAT	TGATGAGTT	TATTATTTA	ATTGAATT	TTTTTTTT	GATGAAGATA	1020
GTTGGTGGTC	GAGTGCCTA	AAGAAGTTAG	AAGGAATTAG	AATTTAGT	TTTATATTT	1080
ATTATTAGAT	ATATTATAT	TTATATACGT	TTTAGATAT	ATATAAGAGT	GTTGTCGGT	1140
TATATTAAAT	TTTATTATTA	TTGTTGTAG	AAATTAAATT	AAAAAAATAA	TAATAATAAT	1200
AAATAATT	AAAAGGATA	AAAAAATTAA	TGATTGAGAA	AAGAGGTATT	TTTTTTGAT	1260
ATTTGGTTT	GTTCGAAATA	ATAAAAGAAG	AAGAAAATT	TATTATTATT	ATCGATT	1320
TTGTTTTT	TTTTTTTT	TTTATT	TTGAAAATC	GTGGGTTGG	GATTGTGAAT	1380
TATTGTATGA	TATTA	AAAAAA	ATAAAAAAA	GTTGAATTAA	AGGGTTTTG	1440
GATAGGAGTG	GT	TTTGT	T			1461

ANGABEN ZU SEQ ID-NO:10:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2536 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:10:

AGTAGTAGTA	GTAATCGGGT	CGTCGTCGTC	GCGGGGTTAA	GAGATTACGT	TTAGTAAAAA	60
TTTTTTAGG	AAAGAGGTTC	GAGAGTTTT	TTTTATTTC	GTAAAGGGTG	TTGATTTGT	120
CGCGTTACGT	TACGTATTAA	AAGGTTCGCG	AAGAAAAACG	CGGGCGAAGT	TTAGTATT	180
CGAATAGAAA	GGGAATTAT	TTAGAAGGG	GGGTTACGG	GTTGAGGGT	TTTGGGTGG	240
GGAAAGTAGT	ATTGGAGGGA	GAAGATAGAA	GGGATGGTAA	GATAGAGGAA	ATTACGGGGA	300
AAAAAAGGAA	GGGATGAGGA	GT	TATAAAAGGG	ATTGACGGAC	GGGATGTAGA	360
GTAATGGTT	TTTTCGAAA	AATGGATTG	TAGGTTTTT	CGTATT	AATAATT	420
TATTATTGA	AAGCGGGGAG	TTGAGTTG	GGGTTCGTA	GATTAAAGTA	GGGTAAAGACG	480
AAGGGAAAGA	GGGAGGGGAG	TTGGAAGTA	AAGTGAATAA	AGAGGTAAA	GAGGTAAAGT	540
TGGAAATGGA	AACGAGAAAG	GGAGTTGGAA	AGGGTCGTT	TTGGAAGGGT	TTGAAAAAG	600
GT	TTAGAAAT	GGGATGTTGG	GGAGGTAAAG	GGGGATAGTT	TCGAGGAAGG	660
AATGTAAGGG	AGCGTTTAG	GT	TTTGGGAA	GAGGAAGGG	TGGGAGGGCG	720
CGAGGTTCGG	GAGTTGGAAT	AGAGGGGTT	TTGCGCGGGG	TTAGGAGGG	GAGAGTAGGG	780
TCGGAGGGAG	GCGATTATCG	CGGGAAGAGT	TAGAGGATGG	GAGGGAGGG	TGTTGCGGT	840
GAAGAGGGGA	TGTGGATT	TTAATGTTA	GGGAGAAGAG	GGAGGTGAAG	CGAGATAAAA	900

GGAAAGGGGC	GAATAGGGAG	AAGAGGAAGG	AGTTTGCAC	GCGATGGGG	AGGGGAGGGA	960
GGTCGTGGGA	AGCGGTCGGG	GGCGCGGGG	ATGGGAAGGG	GTCGGCGGG	CGGCGTGGTT	1020
ATTTAGGTT	CGGGATT	TTTTGGGGA	GC GGCGTGT	CGGC	GGCG	1080
TTTCGGTT	TTCGTTCGT	TTTCGTT	TTTCGGCG	GGCG	GGCG	1140
ACGGTTATA	CGGGCGGTT	TTTGCGTCG	AGTTTCGGAT	GTTGTTTTG	GGGAGGCGGA	1200
GGTAGCGGTA	CGGGTAGCGG	TTCGGTCGT	ACGGTTATTA	TCGTCGCGG	TAGTAGGC	1260
TCGCGGTCG	AGTTTTAG	TAGCGGCGT	CGGC	GGCG	GGCG	1320
TGGTTTTG	CGTTGTTT	TTCGTTTC	GGCGTTTG	GGCG	GGCG	1380
CGTATTTCG	GTTCGCGC	GGTTAGTTA	GCGCGGCGC	TTTTTCG	GCGCGTC	1440
GTTTATTAT	AATCGCGTTC	CGGGAGTGTT	TTTCGTATCG	TTTCGGTG	GTGGGCGGGA	1500
GATATAAGTT	TAGTTAAGTT	TATATAGTTC	GGTCGGGTG	GTTATGGAGA	GGCGTATTGT	1560
AGAGATATT	GGACGCGTAT	TGTGATAGGC	GTTATTTGA	TATACGTGTT	TTTTTATT	1620
TAAATATTAC	GGGCGTAAG	CGTTTATATT	TATATTATT	TGTAGGGATT	ACGTGGGAGA	1680
ATTTGAAGGG	TTGATTGTTG	TTCGGTGTAT	TCGTAGTAGT	AATGATGCGA	GAATTGTT	1740
TAATTCGTT	ATTGATTGGA	TTAGATTGTA	AATTTTG	CGTTGGGAT	GGTTAGTCG	1800
TGTTTAGTAT	TAAGTGTAGG	TATTAATT	TTTGTGGAGT	AAAGTATTGG	ATGAATGAAT	1860
GTATTGGAAT	TGTTATTGA	GTATTGTAA	TAATTTGAT	TTTATGGTT	GGTATAAAGA	1920
ATTTTATTGT	AGTATGAAAT	AAATGGTGT	TTGTGTGTT	TTTTGTTT	GGTTAAGGT	1980
TTAAAATAGG	TTAATGTATG	TTTTTTTT	TTAACGTAG	AGGAGTTTA	TTTGT	2040
GTGTTAGTT	TTATTATAGA	ATGCGTAATT	TTATTTAAT	TTAAATGTT	ATATTTTTA	2100
GGGTAGGAAT	TGTTGTTAG	AATATATAGT	TTGAAGTATA	TTGTGTGTT	ATATAACGAT	2160
AGTTTTTG	TTTGGTAAG	ATTTTTATT	TTTATTG	TTGTTATT	GTTTATTAC	2220
GATTTTTGT	AATTTTATT	TTTATT	TTGTTTTA	TTATTAT	AATTTTTA	2280
TTGTTTTTG	TTTTTTTT	CGTTAGTTT	TGGGGAGTTA	GTTGTTATGA	GTATTTAGTT	2340
TGGTTTTAGA	TGTTGAAAGG	GTTAGTGTAT	ATAGAAGATA	TATATTAGT	GATGGGTAAG	2400
ATGTTGTAAT	TATATTGGGT	TATGTTGAAA	TTGTGAAGTT	TTAATT	TATGTGTAAG	2460
AGAAATTAAT	GTATTGTAAT	TTCGATGTT	GTGTTGAGGA	TTTTTTG	AAATTAAAGT	2520
TTTGGTTT	TGTTT					2536

ANGABEN ZU SEQ ID-NO:11:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 504 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:11:

AGTAGTAGTA	GTAGCGCGTT	GAGTCGGTT	ACGTAGAGCG	TTAGTCGGT	GCGGTTTGT	60
AGGTGGCGC	GCGTGAGTAG	CGGTAGCGC	AGTATCGAGG	CGTTAGTAG	GAAGGTGGCG	120
AGGTTGAGAC	GCGTTTACA	GTTCGGCGGA	GTTCGCGTCG	CGTAGTCGT	TATGGCGTCG	180
GGCGCGGAT	TAGCGGGCGG	CGCGTCGGT	TGTAGTCGG	AGAAACCGT	CGGTTGCGTT	240
TGCGTATTGT	GTCGTGATG	TCGGTTCGGG	AAGGGTAGTT	TTTCGTAGTT	GCGCGTCGTT	300
TGTTAGTTT	CGCGAGAATT	TCGGTTCGTT	TTGTTGTTGT	TCGGAATTG	GCGGGAGCGT	360
TTCGTTCGTT	CGTTTTTT	CGTTTTCGG	GGATATTG	TTTGAGTT	AGATTTGTG	420
TCGGCGGGGG	GTCGGGAAGT	GGTGGGAGAA	GTCGTGTC	GCGTTGTT	AAATTAGTT	480
TTAAATTAG	GAGGTTTTG	TTT				504

ANGABEN ZU SEQ ID-NO:12:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2036 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:12:

AGTAGTAGTA	GTTTTAATT	GAGTTAGCG	TATTCGAAGT	TTCGTTACG	TTTTTTTT	60
TAGAGATT	TTAAGTCGGAG	GAAAATAGCG	TAGTGTGGT	TTTTAGTTA	GTCGTTGTT	120
TTTCGGTTT	TAACGGTCG	GATAAGATGT	AGATGGAATT	ATCGTTTG	TTAGGTTGA	180
GTTTTTATT	GTTTTTAG	TAGTCGTCGT	CGTTTAGGA	GTTCGCGGTA	TCGGGCGCGT	240
CGTTGTCGTC	GTTTTCGGT	AGTATTGGT	TTACGGGTAT	TATTAACGCCG	GTAGAGGATA	300
GTTTTTTA	GGGGATTATT	TTAGTTAACG	GGATTATGTT	TTTTAGAAAT	TTTCGTTATT	360
ATGTTAATT	AGTTTCGGA	GGTATT	TTTCGTAGAT	CGGTTGGCG	TAGATTTAGT	420
ATTATTAGTA	GTCGTCGTC	TTTGCCTCG	CGTCGTAGTC	GGTATAGTTA	GCGTAGTTAT	480
TATAGGCGTA	GTTTCTGTAG	TAGCGTCGTT	TATTCGTTAG	TTTTAGTTAG	GCGTTTACG	540
CGTAGAGGAG	CGTCGTCGCG	GCGTACGGTT	ATTAGTTAT	TATGATTAGT	AAGTCGTTT	600
CGTTTCGGC	GGTTGTAGTC	GTTGTTGTCG	TAGTCGTCGT	TTCGTCGTT	TCGTTAGTT	660
GGAATACGTA	TTAAAGCGTG	AATGTAGTT	GGAGCGTATC	GTAAATT	TGGGGCGGTT	720
TGTAGGCGGG	TCGGGATT	CGTCGGCGG	TCGGTGTGGG	C GTGGGTGTG	GGTGTGGGG	780
TGTTTTTTC	GTTAATT	ATTCGTCGT	TTAAAAAGTT	TTTTTTAGT	AACGTGATCG	840
CGTCGTTAA	GTTTTCGC	GC GGTTTTT	TTATTTAA	GTTTGGATG	GAGGATAACG	900
TTTTTCGGAT	CGATAATGGT	AATAATTGT	TGTTATTTA	GGTAATATT	TGTTGTATT	960
ATATATTTT	TTTTTTGTT	TCGTTTTT	TGTTATTTT	TTGATTTA	ATTATTTATT	1020
CGTTTATATT	TTAAAAAGGGT	AGTAATGTT	AGTTTTTT	TTTTTTTCG	AAGTTTTAG	1080
TTTTTAGGG	TTGTTTATT	ATTAGAGGAT	GAGGTTGGGA	GAATATTGTG	ATTATTGGAG	1140
GAATCGTATT	CGTTTTTA	GGGTAGAAGA	AGTTTTTT	TTTTAGTTA	TTTTTTTTT	1200
TATGGTGT	GTATTTTA	TTTGATTT	TTTGATAAG	TAAAGTTGTA	AGTGTGTGGT	1260
AAGGTGTCGG	TATAGTTA	GGATGAGTAG	GTTGAGATT	TTATTATT	GAGTAGTTAT	1320
ATATTTAGG	TTATGTAATT	TTGAGTTAG	GGTGTGTT	GTAAGCGGTT	TATATTAA	1380
TTATTTGAT	TTTATTTA	AAGAGGTTAA	TAATTTGA	GTGGTGTAT	TTAGACGGT	1440
TGCGTATGTT	TAGTTAAATT	AATGTGAATA	TATGTGTTA	TGTTCGTTA	AGGTGTTAGA	1500
ATTATTAATA	ATAATTAGTA	TATTGTTT	GTTGGGAAA	TTATGAGTGT	GAGATTTAA	1560
TAAATATTA	TTATTGTGTA	AGGGAAAGGAG	GTGGAAGAGT	GGAAATT	AGGGTAATT	1620
TTTTATGTT	ATTAGAAGA	AGGATTTT	TTTTTTTT	CGGAGGTAAG	AGATAATAGG	1680
ACGTGATT	TAGAGTTATA	TTGTAATGGA	GTTTATTGT	TAGTTAGTAT	TGAAATATT	1740
AAAGTTGTAG	TGTTTTTAT	TAGTTGGTT	AAGAAAATAT	AGATTAATTA	TATTATTGA	1800
AATTGGGTAA	AGTGAAGTT	TTGTAATT	AATGAAAAT	GAGTTTG	TTTGGTGT	1860
AAATGTTAAT	AAAAAATT	GGATGTGTT	AGATAAAAAA	GGTGAAATG	AAGAGTGT	1920
TTATTTTGT	TTTGTTTT	TGATTTAGT	TATAGTTATA	GTAATTGGTA	GTGTTTGGG	1980
TATTTGGGT	TGGAAAGGCG	TAGTGCGGGT	GTGTGTTAT	AGGCGGTTT	TGTTT	2036

ANGABEN ZU SEQ ID-NO:13:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 452 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:13:

AGTAGTAGTA	GTAATTTTT	TGTATTTGTA	GTAGAAAATT	TATTTAGAG	ATTGGGTATG	60
TGTGTTAAGG	AGATGGTGAT	TTTTTTGGA	TATTTAGAG	ATTTTTTTT	TTGTTATGAT	120
TTTTATTGCG	GGGATATT	TATAATGTTA	TTTTTTGTA	TTAGTTTG	TTATATTAA	180
ATTGTTATT	TTTAGTTG	TTGTTGTT	TTAGATTATG	ATATTATT	AGTTGTTAG	240
TTTAGGAAAG	ATTTTGGGT	GTATAAATT	GTTAGAATA	GTTGGTAGTA	ATTTGTAGAT	300
CGTTGTTGT	GTCGTGAGGG	TATGGATGTG	GTATATT	ATGGTTAT	TTATATTG	360
GGGGGGACGAG	ATTTTATT	TGGAGTTAAG	TTGAAGGAAG	TGGAATGTTA	TAGTGTGTTAG	420
AGAAATTAGT	GGGTATTGGT	GGTTTTGTT	TT			452

ANGABEN ZU SEQ ID-NO:14:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 513 Basen
ART: Nukleinsäure

STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:14:

AGTAGTAGTA	GTAGCGGAGA	CGGAGGCAGC	GGCGGAGGCG	GCGCGGCCGA	TTTAGCGCG	60
GGCGAGCGG	CGGGGTTAT	GGTCGGATT	GGCGAGTTGG	TTAATAGTG	GGGTTAGGGG	120
TTTCGGCGC	GGGGGGGATC	GGAGGGAGCG	AGGTCGTTGT	CGGACGGGCC	GGGGCGTCGG	180
GAGGGGGCGG	GTCGCGGTG	GACGGGGCGG	GGTATTAGGA	GGAACGGAGT	GGCGTGG	240
CGGTTTCGC	GTAGAGATTG	GGTCGCGGGC	GTAGAGATTG	GGTCGCGGGAG	AGTTAGGTG	300
TTCGCGGTTG	AGGGAGTTGG	AGAGGGGAA	AATCGGGGTC	GTAGCGGGAA	GGCGGAAAGT	360
TAGAGAGAGG	AGGTGTTTCG	GTAGCGGACG	CGTTAGCGGG	GTAGATAGGA	AAATAGTTGG	420
AAAGGGCGAT	TTGGGGGAGT	AGAGATACGA	TTAGAGTTG	AGAAGGGTAG	GAGTAAGGGA	480
TGTTAAGGG	GTAGAGTTG	AGGTTTTGT	TTT			513

ANGABEN ZU SEQ ID-NO:15:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 980 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:15:

AGTAGTAGTA	GTATTTTCGG	TTATAAGGAT	TTTCGAGTT	TCGTTCGTCG	GTCGCGGATT	60
CGGTTTTTT	TTTTTCGGTC	GTAGGGGGC	GGGTTCGGAT	TATAGGATTG	GAGTTGGCG	120
GAGATTACG	TTCGGAGCGG	TTGTGAATTG	GTAGGCAGTG	GGCGCGGTT	TTGTGTCGTG	180
TTTCGGGTAT	TTAGTTTTT	AACGGGTTT	CGGAGTCGAA	GATAGTTTA	GGGTTAGGG	240
AGCGCGGGCG	GTAGGGGGC	GGCGTTAGAT	TGCGGTGAGT	TGGTCGGCGT	GGGTTATTAA	300
TTTAATGTAG	TTAGGGCGG	CGGTACGAGA	TAGAATAACG	GCGAATAGGA	GTAGGGAAAG	360
CGTTTCGAT	AGGTTAGGTT	TAGGGATTG	CGGGGAGAGG	GCGAGGTTAA	TATTCGGTAT	420
GGGTTTTGA	TTGGTTTTG	GGATTGTTT	CGTTACGTT	TATAGGTGGG	TCGTATT	480
TTTTGCGTT	TCGTTTCGT	TTAATAGTT	TATAGTTGTT	GTAGTTTAT	TCGTACGTT	540
CGAATTTCGT	TCGAATTTCGT	TATTGGTTGT	TTTTAGCGG	TTGTGTTGA	TTGGTTGTT	600
GAAGATTTCGT	TTTTTGTG	TGGGTTAGT	TTCGTAATG	CGTAGTTAAG	CGGGTGGTAA	660
GGGGCGGGTG	GAGCGCGGGG	CGCGACGGCG	GAGGGGGCG	TGGGTAGTCG	GACGTATT	720
GGTAGGGAGT	AGTAGGTGGC	GGCGGTGTAT	GGGGTTGGT	TTTATTAGCG	GGTATTGGTT	780
TATAGTTACG	GTCGGGGGGT	TATTTAGTTG	GAGAGAGAAG	GGATAGGTGA	TTCGATCGGA	840
GTAGAGTTA	GTAGAGTTG	GTAGAGTTG	AGATAGCGAG	GGGAATCGAG	GTTGGGGAGG	900
TTATTTAGGG	AGATTCGGA	GGGAATTGG	TGAGGTTGA	ACGGAGGGAG	ATTGGGGTT	960
GAATAAAGGG	TTTTGTTT					980

ANGABEN ZU SEQ ID-NO:16:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 223 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:16:

AGTAGTAGTA	GTAATATTAT	TTTATTAAAA	AAATATTAGC	GTAGTTTTA	ATAGAGTTT	60
ATTTTTAAA	GTGAAGTTAA	AAATAGGTTT	TATGTTAAA	GTAGTTTTTC	GTAGTAGGCG	120
ATACCGTAG	ATGGAGAAAA	TATTTTATT	GTTGGGAAAG	GGAGGCAGTGT	TAACGGGGAC	180

GAAGGATATT TTATTATTTG GTAAAGTTAT TGGTTTTGT TTT

223

ANGABEN ZU SEQ ID-NO:17:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1145 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:17:

AGTAGTAGTA	GTTTGGGTT	TCGTTAGTC	GTAAGATGAG	TTAGTTGGGT	TTGGTGGGG	60
TTGTATTTAG	GTTACGTTT	TTTAGATTAG	TTTTTTTTA	TTAGATTACG	TTATCGTT	120
TTTAGTCGT	TTTGTTTA	GGTTTTTTT	TGTTTTTTC	GTTTGAATT	TTGTTTCGAG	180
TTTCGTTAT	TTTTTCGGT	TTTTTTAGT	TTGTTTTTC	GGATTTGGT	CGAGTATT	240
AAGTTTTT	TTTGGCGATT	AGGTTTTTT	ATTATTTAGG	GTCGTTTTT	GTTGTTCGT	300
TTGGGGCGGT	TTGTTAGTTT	ATCGTCGAGG	AAGAGGAAGG	CGTCGTGGT	TTGGTTTCGT	360
AGCGATTCGT	TTTCGTCGTA	TCGAATTCGT	AGTTTATTA	TTATGTTAT	TACGTCGATA	420
TTTATCGTTA	GTATTATGTA	CGTTATAGT	ATCGTTAGCG	TGGTTGGGA	GAGCAGGCGT	480
ATTAGTTTC	GCGGGCGTCG	TTTAGTTAT	TACGGGTTCG	GTCGTAATCG	GAAGTTTTT	540
CGTTTTTGT	TTTAATTAG	GTGTGTGGT	AGGAGTCGAG	TTTATCGAT	CGGGAGTTAG	600
GCGTTGGAT	TGAGATTGG	AGGTTATATT	AGGATTGCG	GTATTGTCG	TTATTAGGAC	660
GGGGAGGGGG	TCGGGGATAA	TTTATAAGT	CGAGGGCGG	GGTCGTGGC	GGTTTCGGTC	720
GATAAGGGAT	AGTTGTAATT	GAGTCGGGG	TGGAAGGGGT	GGCGTTTCG	TTATGGGTT	780
AGTTGTTGT	TTCGAGTTT	GGGGGTTTGA	AATGTTTTT	GATATTGGG	CGGGCGGAGT	840
TTTATTTGG	TAGGGGATGT	TGGAGTAGAT	GTCGTGGTG	TATTTTGTT	ATAGGTTATT	900
GTGGTTTTT	TGTTGGCGGG	TTAGTAGTT	GGTGGTCGTG	GAGAGTATTA	TGAGGACGTT	960
GGTTACGAGG	TTGAGTAGAA	TGTTTTATT	TTGGAGGTTT	CGTTTATTT	TTATGTTATT	1020
GAGGTTGTAG	TCGGGGGTTA	TAAGGGTAGG	ATGGGTTTAG	GTTTTTTTA	TTTCGAGGTT	1080
TTTAGTTGAG	GAGTCGTTT	TTTTATATT	TTTTTTTGG	TATTTTGTT	TGAGGTTTT	1140
						1145

ANGABEN ZU SEQ ID-NO:18:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 633 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:18:

AGTAGTAGTA	GTATAGTCG	GCGTTGGTTA	GCGGTTAGA	GTCGGGGTAG	GGAGGGTTT	60
CGTGCAGGACG	GGGGTTAAGA	GGTCGTTTT	AGTCGGGGT	AGTCGCGTA	TTCGAGTTG	120
GAGTTCGGAG	TTGGGAGGGG	TGGGGAGGTT	TCGCGTTAG	GGTCGGATCG	GAGGGAGAGG	180
GAGAGCGGTG	TTTTTTTTT	CGGTTTCGT	CGTTAGTCGA	TCGGGGCGTT	GGCGGGCGT	240
CGCGGGAGTC	GTAGTTTTT	TCGGGGGGCG	GATTCGGTT	CGGCAGCGGT	ATTAGTTTT	300
TGTTTCGATA	GGTGCAGTC	GCGCGAAGGA	ATGTAGTCGG	TTGATTAT	TAGCGTTTT	360
TTTTTATTG	CGCGTTCGT	TTATAAAGCG	TTGTTCGTT	CGTTTTATT	TTTTAATT	420
TCGCGTTCGT	TTTCGGATAG	TTTTGTTCG	TTCGCGCGT	GTAGTTTAT	TTTTAGCGG	480
TAGTTAGGC	GCGGAGGGAG	CGAGTCGTT	TCGAGGTAGG	TTAGGACGG	GCGTATAGTA	540
GTAGTCGAGG	TTGGTCGGGA	GAGGGTGAGT	GTTCGTTTA	GTTGGGTAGT	CGCGTAGGGG	600
AGTTGGTCGG	GTAGGGTCGG	TGGTTTTGT	TTT			633

ANGABEN ZU SEQ ID-NO:19:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:19:

AGTAGTAGTA GT

12

ANGABEN ZU SEQ ID-NO:20:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:20:

ACAAAAAACTA AA

12

ANGABEN ZU SEQ ID-NO:21:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 74 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:21:

AGTAGTAGTA GTTTCGGTAG TTTAGTTAT GGCGGCGGTG CGGGCGGTAG TAGGTTGAG
TTTTAGTTT TGTT

60

74

ANGABEN ZU SEQ ID-NO:22:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 103 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:22:

AGTAGTAGTA GTAGCGGTAG CGGTAATAGG GCGGTTGAGA ATTCCGGCGGC GGCGTTTTT
TTCGTTTTT TTTTTTCGT TTCGTGATT TTTAGTTTT GTT

60

103

ANGABEN ZU SEQ ID-NO:23:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 559 Basen
ART: Nukleinsäure
STRANGFORM: Einzelstrang
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:23:

AGTAGTAGTA	GTAAGAAGGA	AAAAAAATAAA	TTTTTTAAGT	ATTTAGTGTA	GT	TTTTTTTG	60
GGTTTTTAG	TTTTATTTCG	GTATAGTTAG	GGTTTTAAT	TTATTTTGT	TT	TGGGTTA	120
GGGAAAATAA	AAAGTTAAGT	TTATTCGTAT	TTAGTTAAA	GTTGAAAATA	AGTGGTCGGA	180	
TTAATTTCG	TTTTAAATA	AATTATTGA	TGGTGTAGA	AGGTAAATAAA	TTTTATT	TT	240
TTTTTAGAT	AAGGGTTAT	TTTACGAGAT	TATTAAGTAT	TGTTTCGTGA	GATTATTGAA	300	
ATAATAGTAT	TTTTTTTTT	ATCGAGGATT	ATAGTATGTT	TTTGATTAA	AACGTTTAA	360	
GT	TTTTGAAG	AGATATTGG	GAGTTTTAG	AAATCAAAGG	AAGGATTGGA	AATGTTAGGG	420
GGTGAGGGC	GTAAATAAT	TTTGATT	ATGGTTGTGA	TTGATGTTGT	TATTAGATTA	480	
TTTTATATT	TAATAATATT	GAATTATTAA	AACGGATTAG	TATAAATTAA	ATTATGATAG	540	
TATATTAA	GT	TTTGTT				559	

ANGABEN ZU SEQ ID-NO:24:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1695 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:24:

AGTAGTAGTA	TTTTGGAGA	GTAAGTGGGT	TTTATGAAA	GTGTTATTG	TTATATATTG	60
TTATTTTAN	AGGATTTAG	AATATTTCG	AATTGAAAAA	TAGAGTGAGT	GTGGAGGGGA	120
GGGGGGGTGTT	TACGTGGAGG	AGGGTCGCGA	GAAGGAGAGT	GCGTCGGGTT	GATGGTTAGC	180
GGTGTTCGG	GGGTGATGT	CGGGCGGGAT	CGGCGGCGCG	GGGGTGGGGC	GACGGCGGGG	240
CGGCCTGCA	GTAGTAGAGG	GGCGTTCGT	AGAGGTCGGG	GGGGGCGCGG	GGTTTCGCGT	300
CGTTTGT	GTAGTCGTT	CGTCGTTTT	TTCGGTTCGT	TCGTAATCGT	CGCGGGATT	360
GTCGGTCGG	GT	TTTGAA	TTTTTACGG	GGGATAGTCG	ATCGGGGGCG	420
TTTCGTATT	CGATTCGGAC	GGCGGTAGTA	GGGGGAGGGA	TGGTGTATCG	AGGAAGTTAA	480
GGTTTATTA	TTTGTTCGG	CGGGTCGGCG	GTTCGTTGGT	TCGTTGGTC	GGAGAAGTGT	540
TGCGTTAGG	TTGGTCGTA	GGAAACGGCG	GCGCGGTT	ANTNTANTT	TTNNNNNTNT	600
TNNNTGNAN	TNTTTTTA	TAGGGGAGGG	GTAGGCGGTT	CGCGGGTTTC	GCGTCGGTCG	660
GGCGATTGGG	TACGCGAGGG	AGCGGGTAGG	GTAGGGGGAA	ATAAATTAAG	GTCGAGATT	720
AGAGTCGGAG	AGCGCGGGAG	GAGTAGCGGC	GAGAGGTAGG	AGAGGTAGAG	AAGAAGAAAG	780
GAGGGAGAGA	GGGGCGAAG	ACGGTCGGGA	AATTCGGGGT	TGTAGTTTCG	CGTCGCGTCG	840
GAGTCGTGAC	GGATTATTG	TAGCGGTGCG	GTTGATTCGT	AATTTAAGG	TACAAAAAAA	900
GGGGAGGGTA	GAGAGGGAGG	GGAAAGCGGA	GTGTTGAGC	GTCGGGGCGG	GGCGGGTT	960
TCGCGAGCGT	CGTATATTG	GGAATTGTN	GTTCGTTGC	GGCGAGTGT	CGCGTGT	1020
GGCGAGTTT	GGTTGGGAA	NTTGAGCGC	GCGGATTACG	TTCGTTTTT	AGTCGGTCGT	1080
TTTTGCGTG	TAGGGTGGGG	GGAGGTTAGG	GGCGGGGGCG	CGGACGTCGG	TGACGTCGCG	1140
TGGCGGAGTT	TTTCGGTTA	TGTGGTTGGA	GGCGGGCGGG	GAAGAAGGTA	AGGTTGGGAG	1200
GGGAAGGAGG	AAATGCGAGG	GT	TTTCGGGC	GGGAGGAAGC	GCGAGGGGTG	1260
AGCGAAAGGG	GGATAGGGGT	TTGGAGTGG	AAGTTTGCG	GGAGAGGAGG	CGCGAGTGT	1320
TAGTTTGT	GTGTTAGAAT	TATTGGATG	TCGTATTAAA	AAATAGGAA	AATTAAAGTT	1380
TTGCGAGAGG	TAGTTAAAGG	TATTGTTGTA	TTCGTGTATC	GATTGGATT	TATTATAGAA	1440
TTGTACGGTT	TATTAGGATT	GTGTTATTT	CGGTGTAGTA	TTCGTAGGTT	TATTGTAAT	1500
ATGAATTGGT	AGAGTTAGGA	ATTAGGTTGT	AAAGATAGAG	AAGCGAGTTA	AAGTTGGTCG	1560
TAGTCGAGG	CGTGGGGAGA	ATTGGGTAA	ACGAGGAGAA	GGGATATT	TTATTGTTAG	1620
AAAGATT	TATTAGTT	TGTATTTA	TAAATCGTT	AGATT	TTGGCGGAGT	1680
TTATTAGTT	TTGTT					1695

ANGABEN ZU SEQ ID-NO:25:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 722 Basen

ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:25:

AGTAGTAGTA	GTTTCGTTAC	GAATACGATT	AGTTTATATT	GGTTGTTAG	TATTCGGTT	60
AGTTTTGTG	GGACGTTGGC	GCGTATATCG	TTTCGTCGGG	GGGTCGGGGA	TTTCGGGGAT	120
TCGGGGTCG	AGGACGAGGG	AGGCGAGTAG	GTCGTTGGTT	TCGACGATT	TGCGATCGAG	180
GCGTTTGAG	TTGTTCGGGG	CGAGGCGGGC	GATTTTAGGT	TTTCGAGTAT	TTTCGAGATA	240
TTCGCCTAA	GTTTGGTTTC	GGGGGTGGTT	GTCGTCGGGT	TTCGTTGCG	TTTTGGTTT	300
TTGTTTTGT	TTCGACGTTT	TTGTTAATTA	GTATTTTTT	TTTAATTTT	TTTCGATT	360
TTATTTTACG	TTTTGTTAA	TTGTTTTTT	TTTTGTTAT	TTTCGACGT	TCGTTTTT	420
TTTTTTGTT	TTTCGTTTT	TCGTTAACGGT	ATTATTTGT	TTATTTATT	AGCGTTTAT	480
TTTGTGATT	TGGGATTTA	CGAGTTTTT	TGTTCTGTTGT	TTTTTATTGG	GTAACGTCG	540
GGGTAGTTAT	TTGTTTTTT	CGGGATTAC	GCGGATAGTT	TTTCGTTTT	GATTTTGGG	600
ATATTGGTTA	GTTTGTGCG	GATATTAGCG	CGTNNTTCG	TATTTTCGTT	CGCGCGTCG	660
GTTTCGTT	GTCGTTGTA	GTTTTATT	TTGAGTCGAC	GTTCGTTAT	TTAGTTTG	720
TT						722

ANGABEN ZU SEQ ID-NO:26:

SEQUENZCHARAKTERISTIKA:
 LÄNGE: 517 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:26:

AGTAGTAGTA	GTTCGAATT	GCGCGCGTAG	CGGCGTATGG	TTAGGTTAG	CGAACGTAGT	60
TTAGCGAGT	GGCGCGTTAG	GCGTATTACG	TAGAGTACGC	GTAGCGTAGC	TAGTAGTCGT	120
AGTATTAGTT	TCGCGCGTT	TAGGAGTTG	GTTCGTTCG	GGTTTGTGCGT	TAGTTTAGT	180
AGTAGCGATA	CGTAGAACGG	TAGGAGCGTT	AGGATGTTAA	TGATGTTGAG	TGGCGCGCGT	240
AGGAAGGC	ATTGTTTTC	GGTTTGTAGG	GAGCGTAGTA	GGAATTGAA	GGAGAATTAG	300
GTTACGTATA	CGGTTTTAG	TACGAATAGG	TTGCGGTATT	TGGGGGAGTA	TTCGTTTGC	360
GGGGAGAGGG	GATACGGGGT	TGGGGCGTA	GGTTTTTGG	AGGGTCGGGG	GCGTTTGTT	420
TTTTTTTTT	TAGGATTAGA	GGCGTTTTT	AGGTTATT	TTTCGTTTA	ATTACGGAA	480
GCGGGTG	GGGATATTGA	TAGGTTAGT	TTTGTT			517

ANGABEN ZU SEQ ID-NO:27:

SEQUENZCHARAKTERISTIKA:
 LÄNGE: 1078 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:27:

AGTAGTAGTA	GTATTAGTAA	TAGTAGTACG	AAAAGTAAAA	TTGTAATT	AACGGTTTT	60
AGGGTTAAG	TAGGTTGAC	GAAGATT	CGTT	TTAGAAAAT	TGGGAGTTTC	120
GTTCGTTTT	TCGGTATTGA	AACGCGATCG	GT	TTTGTGTTG	GTATCGTATT	180
TTTTATTAT	ATTACGACGA	CGGACGTTCG	AGAACGTTGT	ATCGCGTTTC	GTAGGAAGTG	240
TTTTTTGGG	CGGAAGTTT	TGAGCGTGAT	ATAGCGGAAG	TGTTTTTTT	TTCGGTTTT	300
TTGGTTTCGG	TCGTAGAAGC	GAGATGGTGA	GTTGTGATTG	TGGTGTGTTGT	GAATCGCGTT	360

TTATTTCGT	TTTTGTGTT	TTTTGTTTG	TTGTGTTGG	GGGGTTGGTA	AGATTCGGA	420
TAAGGGAAT	TGGTGGGTTG	GAAAGAGGTA	TGCGGTGGTT	TTAAGAGTT	AGAAGAATGA	480
TTGTTAATTG	GTGTTGGGG	GATTTATTTC	GTCGTAATTG	TGGTGTAGA	GTCGTATTGT	540
GTTTTGTT	TCGGTTAAC	TTTGGAGAT	TTTACGGT	TTAGTTTG	GTTGGGAGTC	600
GAGGGAAGGA	GTGGGAAT	GTGGTTTT	GTGTAATAAT	GAAATAATTT	ATTGGTGATG	660
TTTTTGGTC	GGAGTTGTA	AAGATAAGGT	GTATTTAGA	ATATTGTAAT	TTTGCGGAG	720
GGTTTAGGT	AACGTGAAAT	GCGGGTAGTG	GTGTTGATT	TGGTATTGTT	GGAAAGAGGT	780
TTTTTCGTA	GTTCGATT	TTACGATTGT	TTTAAATT	TATTAGTAAT	TGGTTTTCG	840
GAGAATTCGA	GTAAATTAG	AAGTTGTTAG	GTGTTAGAAT	TATTATTTT	TTAATGTGT	900
AGACGAAGGG	AACGTTATCG	TTGGAAAGC	GTCGTAATAA	GACGTATACG	TTGTGTCGTC	960
GTTGTGGTT	TAAGGTTAT	TATTTTAA	AGTCGATTG	TGGTAAATGT	GGTTATTTG	1020
TTAAGCGTAA	GAGAAAGTGT	AAGTAATATT	TTTAGGTTA	ATTGTGTTAG	TTTTGTT	1078

ANGABEN ZU SEQ ID-NO:28:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2949 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:28:

AGTAGTAGTA	GTCGTAGGAG	TAGCGTTTG	GGGAGGGGGG	TTCGTTTTC	GGGGTGC GTT	60
TCGTGGTTAT	TTTTTTAAN	GGGGGTCGTG	GTCGTTTTT	AGTTTCGGCG	CGTTGGGTT	120
GTTTAGCGG	GGTCGTCGG	GCGGGGCGGG	GATTTGTATT	C GGGGCGTT	TTTCGGGACG	180
CGCGGGTTT	CGTTTTTTT	TTGTTTTCG	GGCGAGTTG	GATTTGTT	GGTCGCGGGG	240
TTTCGTTAT	TTGTTAAGTG	AAGGTTTAC	GGGAGATAAT	AAAGAAAGAA	GTTGTTTTT	300
TTTTTTTAG	TTAAATTACG	GATGATTAGT	AATCGGTTG	TTTTTTTTT	TGTCGTTTT	360
TTTTTTTTT	TTTTAGGAAT	ATATTATT	CGTTTTTTT	TAGAGTAGTA	420	
AAATTTC	TTTTTGAAAG	CGGTGGTGT	GTTGGTTGTT	CGCGAGTGC	GAGAGGGGAG	480
GTTGGTTTT	TTTTTTTTT	TTGGTTTTT	TCGTTTAAG	TTTGTTTTG	TTTAGGAAA	540
ATTTTTGCG	ATGGTGTATT	TTTTGAAGT	TGTTATTGTC	GGAGTTTTG	GTTGTTTTG	600
AAATAAGAGT	TTTTTTTAT	TTTTTCGTA	ATAGTTATAG	TTGATGCGA	TATTTTTTT	660
TTGGGTAGA	AATTAGGAAA	TAAATATAAT	AAGGAAAAAA	TTGAATAGGA	AAAAAAATGA	720
TGTCGAATCG	TTTGTTATT	TTGTTGTTGA	CGGTTGGGT	TGGTGTGTTA	TTTTTTTTT	780
AGTAGTTGTT	TGGAGAAGGA	GGAGGAAGAA	GAGTGTATTG	GTGGAGGTGG	AGGTGTGTGT	840
GTGTATATAG	GGGATCGATT	TTGATATAT	AATAATTAA	GTGGGTACGT	TGTACGGTTC	900
GTAGAGTGT	TCGACGGCGG	GCGTTGGCGT	TGGTGTGTT	GGAGTAGAGT	TGTAAGTAGGA	960
GTTCGTTTG	CGTTTCGTT	CGCGTTTTT	TTTTTTTTT	TATTTTTTT	TTTTTTTTT	1020
TTTTTTTTT	TTTTTTTCG	GGCGTTCGTT	TTCGTTTCGA	TTGATAGTT	CGTTTTTTGT	1080
TTTTTTTTT	TAGAGCGGGG	ATTGTCGAAT	GTTTATTTT	CGTCGCGAGC	CGTATTTAT	1140
TATTATATT	TTTTAAATAT	AAGGAAGTCG	AGTATTAGGG	TTTCGTTGA	TTGTTATCG	1200
TTGGGTGATT	GGTAGGTTCG	GAATGATTGG	TTAGGACGC	GAGGTTTTT	TTCCGGTCGT	1260
TTCGATTGGT	TGTTTTTAA	TTAGGTAGA	GCGTCGTTAGA	AGTTGGAAAT	TTGTCGTTT	1320
TTAGTTTTT	TTTTTTTAT	TATTTTTT	TTTCGTTT	TTTCGTCG	TTTTTTTCGT	1380
TATATTTAA	TTTTAGTGT	TCGGTTAGA	CGTTGGCGTT	TTTCGGCGG	TTTGGCGTT	1440
CGTAATAGGT	TTGGGCGGGG	GGAAGAAAGG	GGAGATAAAA	GGGAGGGAGG	GACGAGAGGG	1500
GGGAAGAGAA	TTAGAAGGAA	AACGAAGGGG	GAAATATGAA	AAATAGTAAT	TTGTTTATTT	1560
TATTTAGTAC	GCGTCGGGTT	GTTTATTTT	TTATTTCGT	TTCGTTAGAG	ATTGTAAAG	1620
CGCGCGGTAC	GGATGTATTA	TTAAGTTAA	TATTTATAGA	TAGATGTGTT	TTTAGTGGTT	1680
TTAAAGTTT	TTTTTTTTT	TATTAGTCGG	TTTTTTTTT	TGATCGGCGG	GGGTGGGAAG	1740
CGGGGGTTCG	TATCGTTAAG	AGTTTGGTC	GTGGTTACGT	TTTTGGAAA	TTCGTTATCG	1800
TAATTTTCG	TAGTCGTTGC	GCGTCGGCGG	TAGTTTTTT	TGTTTAAGTT	TTCGGATGTT	1860
ATTATTTGGG	AAATTGTAG	TTAGGANNT	TCGGTGTTCG	CGCGTGCAG	ACGGATCGTT	1920
TCGTGTCGTC	GTTTTTGGT	TTTTTATT	TTTTTTTTT	TTTCGTTT	TTTTTTTTT	1980
TTTTTTTTT	TTTTTTTTG	GGATTCGTGT	AAGTAGCGTG	CGTGTGTTTG	TGAGTTGGAA	2040
GGGTGTGGT	TAGCGGAGTC	GTGCGATGTA	TTGAAAAGT	AATTAGTCG	ATTTTTTTT	2100
ATAGAGTTAT	CGTAAGTGT	TTCGATTTA	ATGGTTTAG	AAATTTTGT	GGTATTAGGA	2160
TTCGGTCGAA	AGAACGGGGA	TCGGTTATTC	GCGTGTTTTT	TTGTTTATC	GTGTTTCGT	2220

GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	GTGTGATTAG	CGGTGGGGGG	GTTTGTTCGT	2280
TTTTTATTTT	TAGGTATGGT	GGTGTTCGGT	TTTTTTTTTT	TTTGTTTA	GATGGATTG	2340
CGTGGTGGAT	GGGGTTGGCG	GCGATAAAATG	TTTTTTTAGT	TTAATTAG	TTGAAAGAGT	2400
TAAGGGGGAT	GGGAGGGGGG	GTGTATTCCG	TAGGCAGAG	AAGCGGGGGT	GGGGTGGGGT	2460
GGGGTGGGGG	GACGCGGTTA	AGCGGAAACG	TTGTATAGAG	GAATTTCAGC	GAATTAAGAA	2520
AAAAGAGAAA	GTGGTAACGA	AAATAAGGGT	AAATTGAGTT	TTTTCGGGG	ATTTTAATG	2580
AATTAATTAA	ATTCGGATAT	TTAATAAAATA	TATGGTTTT	AATGAGCGTG	CGTGTGCGTG	2640
TATTCGTAT	TTTAGTTGC	GGGTGCGTTC	TTTGCGTTC	GTCGTTTTA	GTTAGAGTT	2700
GTATTTGGTA	TTTCGAGTT	CGAGCGGTAG	TTTAGGACGT	AGTCGAGGAG	CGTCGTCGGT	2760
GCGTTTCGAT	TAAAATGTGA	ACGGGTTGT	TTTTTTTTT	CGTTGTATCG	TGTTTTGTC	2820
GAGCGTAGTT	AAGTGAATT	AGTCAAAGTA	GGAGTTTTT	TGTTTAGTCG	TAGTTAGGTA	2880
GTTTTCGCGT	GATTTAAGA	TTAATATTAG	ACCGTAGAA	TTTATGTAGG	TTTTGTTA	2940
						2949

ANGABEN ZU SEQ ID-NO:29:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 117 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:29:

AGTAGTAGTA	GTTTATTTG	GGTTGATATA	AAATTGAAAA	ATTTATTGT	TTTTTTATTT	60
TTTTGGGTTG	GATTCGGTGT	ATCGGTTGAT	ATATTTTTT	GGTTATTAGT	TTTGTT	117

ANGABEN ZU SEQ ID-NO:30:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 639 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:30:

AGTAGTAGTA	GTATTATGGT	TATTCGACG	GTCGCGTCG	TTATTTTTT	GCGGCGGTT	60
AGTCGCGACG	TTGTTAGGGT	TAGTAAATT	TTTTTATTTT	TGGGCGTAAT	GGTTGTCGGG	120
GTCGGGGTTT	TTCGCGGGTT	AGGAGGGCGT	GATTCTTTT	GGGTTTGGGT	TTGTATTTT	180
TCGACGGTTG	TTTTTCGTTT	TTTTTATTTT	GTTTTGCCG	TTTTAGTCG	CGGCGGTGTC	240
GTTTTTAAG	TCGTTCTTA	AGGGGAGGTT	TTTCGTTGGT	TACGTTAGGG	GTAGTTTCG	300
ACGGTTTAGA	GTTAGTGGTT	TTTTAGTATT	TTTCGTTTA	GTTTTAACGA	TTTTGGGTAT	360
TTGAGATTCG	CGGTTTTCG	GACGGTTGGT	TTTTTAGGGA	TTTGAGATGT	TTGTTTTTA	420
GATTGTTGTT	TTTTTAGGGA	TGGCGCGGTG	TTTGGGTTT	AGATTGTTA	GATAGATTAT	480
TTTTGATGG	AGAGGGGATT	GTTCGGCGCG	TTTCGGACGT	TTCGGGTTT	GAGTTGCGGG	540
TGTTGTTAT	CGGGCGCGAT	TTTTAGTAG	GTTTGCGTT	TTGTTTTTG	GTAAGTATCG	600
ATTTATTTG	TTATTTTGT	CGCGGTTTA	GTTTTGTT			639

ANGABEN ZU SEQ ID-NO:31:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 304 Basen
 ART: Nukleinsäure
 STRANGFORM: Einzelstrang
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:31:

AGTAGTAGTA	GTATAGTCG	GCGTTGGTTA	GCGGTTAGA	GTCGGGGTAG	GGAGGGTTT	60
CGTGCGGACG	GGGGTTAAGA	GGTCGTTTT	AGTCGGGT	AGTCGCGTA	TTCGAGTTG	120
GAGTCGGAG	TTTGGAGGGG	TGGGGAGGTT	TCGCCTT	GGTCGGATCG	GAGGGAGAGG	180
GAGAGCGGTG	GT	CGGTTTCGT	CGTTAGTCGA	TCGGGGCGTT	GGCAGGGCGT	240
CGCAGGGAGTC	TTT	TCGGGGGGCG	GATTGGTT	CGGCAGGGCGT	ATTTAGTTT	300
TGTT						304